



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 108396

TO: Sheridan Swope
Location: CM1/12S12&10D)1
Art Unit: 1652 12D12 / 10D01
Tuesday, November 25, 2003

Case Serial Number: 09/607745

From: Barb O'Bryen
Location: Biotech-Chem Library
CM1-6A05
Phone: 308-4291

barbara.obryen@uspto.gov

Search Notes

STIC-Biotech/ChemLib

108396

From: Swope, Sheridan
Sent: Monday, November 17, 2003 12:48 PM
To: STIC-Biotech/ChemLib
Subject: 09607745

For 09607745, pls do the following searches and alignments:

Search:

SID 2 against the NT and AA data bases.

SID 9 against the NT and AA data bases.

Align SID 2 and 9 with:

EMBL Acc# AF179224

EMBL Acc# AF216312.

EMBL Acc# Q9NRS4

GENSEQ Acc# AAY99417

CRFF

THANKS!!!

Sheridan Swope, Ph.D.
Patent Examiner, AU 1652
Recombinant Enzymes
sheridan.swope@uspto.gov
703-305-1696 (voice)
703-308-3014 (FAX)
Mailbox: CM1 Rm10D01
Office: CM1 Rm12D12

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 13:51:12 ; Search time 0.001 Seconds
(without alignments)
378.015 Million cell updates/sec

Title: US-09-607-745-2
Perfect score: 2342
Sequence: 1 MDPDSDQPLNSLDVKPLRKP.....VYTKVSAYLNWIYNVWKAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2 seqs, 869 residues

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : aa.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2337	99.8	437	1 q9nrs4	TOIG of: q9nrs4
2	2297.5	98.1	432	1 aay99417	TOIG of: aay9941

ALIGNMENTS

RESULT 1
q9nrs4
; TOIG of: q9nrs4 check: 316 from: 1 to: 437
; ID TMS4_HUMAN STANDARD; PRT; 437 AA.
; AC Q9NRS4; Q9NZA5;
; DT 16-OCT-2001 (Rel. 40, Created)
; DT 16-OCT-2001 (Rel. 40, Last sequence update)
; DT 15-SEP-2003 (Rel. 42, Last annotation update)
; DE Transmembrane protease, serine 4 (EC 3.4.21.-) (Membrane-type serine protease 2) (MT-SP2).
; GN TMPRSS4 OR TMPRSS3.
; OS Homo sapiens (Human).
; OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
; OX NCBI_TaxID=9606;
; RN [1]
; RP SEQUENCE FROM N.A.
; RC TISSUE=Pancreatic carcinoma;
; RX MEDLINE=20283276; PubMed=10825129;
; RA Wallrapp C., Haehnel S., Mueller-Pillasch F., Burghardt B., Iwamura T., Ruthenburger M., Lerch M.M., Adler G., Gress T.M.;
; RT "A novel transmembrane serine protease (TMPRSS3) overexpressed in pancreatic cancer."
; RL Cancer Res. 60:2602-2606 (2000).
; RN [2]

SEQUENCE FROM N.A.
Smeekens S.S., Lorimer D.D., Wang E., Hou J., Linnevers C.;
"MT-SP2, a novel type II membrane serine protease expressed in trachea, colon, and small intestine: identification, cloning, and chromosomal localization."; RT
Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases. RL
[3]
SEQUENCE FROM N.A.
TISSUE=Pancreas; RP
MEDLINE=22388257; PubMed=12477932; RX
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."; RT
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002). RL
-!- FUNCTION: MAY PLAY A ROLE IN TUMOR PROGRESSION, METASTASIS CC
-!- FORMATION AND TUMOR INVASION. CC
-!- SUBCELLULAR LOCATION: Type II membrane protein (Potential). CC
-!- TISSUE SPECIFICITY: HIGH LEVELS IN PANCREATIC, GASTRIC, COLORECTAL CC
AND AMPULLARY CANCER. VERY WEAK EXPRESSION IN NORMAL CC
GASTROINTESTINAL AND UROGENITAL TRACT. CC
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. CC
-!- SIMILARITY: Contains 1 SRCR domain. CC
-!- SIMILARITY: Contains 1 LDL-receptor class A domain. CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF179224; AAF74526.1; -;
EMBL; AF216312; AAF31436.1; -;
EMBL; BC011703; AAH11703.1; -;
HSSP; P00763; 1DPO.
MEROPS; S01.034; -;
Genew; HGNC:11878; TMPRSS4.
MIM; 606565; -;
GO; GO:0016021; C:integral to membrane; NAS.
GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
InterPro; IPR001314; Chymotrypsin.
InterPro; IPR002172; LDL_receptor_A.
InterPro; IPR001254; Ser_protease_Try.
InterPro; IPR001190; Srcr_receptor.
Pfam; PF00057; ldl_recept_a; 1.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SM00192; LDLa; 1.
SMART; SM00202; SR; 1.
SMART; SM00020; Tryp_SPC; 1.
PROSITE; PS01209; LDLRA_1; FALSE NEG.
PROSITE; PS50068; LDLRA_2; FALSE NEG.
PROSITE; PS00420; SRCR_1; FALSE NEG.
PROSITE; PS50287; SRCR_2; 1.
PROSITE; PS50240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.

Perfect score: 1600
Sequence: 1 MDSKGSSQKSRLLLLLVSN.....LNWIYNWKAELSRHHHHH 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2 seqs, 869 residues

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : aa.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	79.1	432	1 aay99417	TOIG of: aay9941
2	1266	79.1	437	1 q9nrs4	TOIG of: q9nrs4

ALIGNMENTS

RESULT 1
aay99417
; TOIG of: aay99417 check: 6343 from: 1 to: 432
; ID AAY99417 standard; Protein; 432 AA.
; XX
; AC AAY99417;
; XX
; DT 08-AUG-2000 (first entry)
; XX
; DE Human PRO1570 (UNQ776) amino acid sequence SEQ ID NO:275.
; XX
; KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis; transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
; KW
; XX Homo sapiens.
; OS
; XX WO200012708-A2.
; PN
; XX
; PD 09-MAR-2000.
; XX
; PF 01-SEP-1999; 99WO-US20111.
; XX
; PR 01-SEP-1998; 98US-0098716.
; PR 01-SEP-1998; 98US-0098749.
; PR 01-SEP-1998; 98US-0098750.
; PR 02-SEP-1998; 98US-0098803.
; PR 02-SEP-1998; 98US-0098821.
; PR 02-SEP-1998; 98US-0098843.
; PR 09-SEP-1998; 98US-0099536.
; PR 09-SEP-1998; 98US-0099596.
; PR 09-SEP-1998; 98US-0099598.
; PR 09-SEP-1998; 98US-0099602.
; PR 09-SEP-1998; 98US-0099642.
; PR 10-SEP-1998; 98US-0099741.
; PR 10-SEP-1998; 98US-0099754.
; PR 10-SEP-1998; 98US-0099763.
; PR 10-SEP-1998; 98US-0099792.

10-SEP-1998; 98US-0099808.
10-SEP-1998; 98US-0099812.
10-SEP-1998; 98US-0099815.
10-SEP-1998; 98US-0099816.
15-SEP-1998; 98US-0100385.
15-SEP-1998; 98US-0100388.
15-SEP-1998; 98US-0100390.
16-SEP-1998; 98US-0100584.
16-SEP-1998; 98US-0100627.
16-SEP-1998; 98US-0100661.
16-SEP-1998; 98US-0100662.
16-SEP-1998; 98US-0100664.
17-SEP-1998; 98US-0100683.
17-SEP-1998; 98US-0100684.
17-SEP-1998; 98US-0100710.
17-SEP-1998; 98US-0100711.
17-SEP-1998; 98US-0100919.
17-SEP-1998; 98US-0100930.
18-SEP-1998; 98US-0100848.
18-SEP-1998; 98US-0100849.
18-SEP-1998; 98US-0101014.
18-SEP-1998; 98US-0101068.
18-SEP-1998; 98US-0101071.
22-SEP-1998; 98US-0101279.
23-SEP-1998; 98US-0101471.
23-SEP-1998; 98US-0101472.
23-SEP-1998; 98US-0101474.
23-SEP-1998; 98US-0101475.
23-SEP-1998; 98US-0101476.
23-SEP-1998; 98US-0101477.
23-SEP-1998; 98US-0101479.
24-SEP-1998; 98US-0101738.
24-SEP-1998; 98US-0101741.
24-SEP-1998; 98US-0101743.
24-SEP-1998; 98US-0101915.
24-SEP-1998; 98US-0101916.
29-SEP-1998; 98US-0102207.
29-SEP-1998; 98US-0102240.
29-SEP-1998; 98US-0102307.
29-SEP-1998; 98US-0102330.
29-SEP-1998; 98US-0102331.
30-SEP-1998; 98US-0102484.
30-SEP-1998; 98US-0102487.
30-SEP-1998; 98US-0102570.
30-SEP-1998; 98US-0102571.
01-OCT-1998; 98US-0102684.
01-OCT-1998; 98US-0102687.
02-OCT-1998; 98US-0102965.
06-OCT-1998; 98US-0103258.
06-OCT-1998; 98US-0103449.
07-OCT-1998; 98US-0103314.
07-OCT-1998; 98US-0103315.
07-OCT-1998; 98US-0103328.
07-OCT-1998; 98US-0103395.
07-OCT-1998; 98US-0103396.
07-OCT-1998; 98US-0103401.
08-OCT-1998; 98US-0103633.
08-OCT-1998; 98US-0103678.
08-OCT-1998; 98US-0103679.
08-OCT-1998; 98US-0103711.
14-OCT-1998; 98US-0104257.
20-OCT-1998; 98US-0104987.
20-OCT-1998; 98US-0105000.
20-OCT-1998; 98US-0105002.
21-OCT-1998; 98US-0105104.
22-OCT-1998; 98US-0105169.
22-OCT-1998; 98US-0105266.
26-OCT-1998; 98US-0105693.
26-OCT-1998; 98US-0105694.
27-OCT-1998; 98US-0105807.
27-OCT-1998; 98US-0105881.
27-OCT-1998; 98US-0105882.
27-OCT-1998; 98US-0106062.

```

; PR 28-OCT-1998; 98US-0106023.
; PR 28-OCT-1998; 98US-0106029.
; PR 28-OCT-1998; 98US-0106030.
; PR 28-OCT-1998; 98US-0106032.
; PR 28-OCT-1998; 98US-0106033.
; PR 28-OCT-1998; 98US-0106178.
; PR 29-OCT-1998; 98US-0106248.
; PR 29-OCT-1998; 98US-0106384.
; PR 29-OCT-1998; 98US-0108500.
; PR 30-OCT-1998; 98US-0106464.
; PR 03-NOV-1998; 98US-0106856.
; PR 03-NOV-1998; 98US-0106902.
; PR 03-NOV-1998; 98US-0106905.
; PR 03-NOV-1998; 98US-0106919.
; PR 03-NOV-1998; 98US-0106932.
; PR 03-NOV-1998; 98US-0106934.
; PR 10-NOV-1998; 98US-0107783.
; PR 17-NOV-1998; 98US-0108775.
; PR 17-NOV-1998; 98US-0108779.
; PR 17-NOV-1998; 98US-0108787.
; PR 17-NOV-1998; 98US-0108788.
; PR 17-NOV-1998; 98US-0108801.
; PR 17-NOV-1998; 98US-0108802.
; PR 17-NOV-1998; 98US-0108806.
; PR 17-NOV-1998; 98US-0108807.
; PR 17-NOV-1998; 98US-0108867.
; PR 17-NOV-1998; 98US-0108925.
; PR 18-NOV-1998; 98US-0108848.
; PR 18-NOV-1998; 98US-0108849.
; PR 18-NOV-1998; 98US-0108850.
; PR 18-NOV-1998; 98US-0108851.
; PR 18-NOV-1998; 98US-0108852.
; PR 18-NOV-1998; 98US-0108858.
; PR 18-NOV-1998; 98US-0108904.
; XX
; PA (GETH ) GENENTECH INC.
; XX
; PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
; XX
; DR WPI; 2000-237871/20.
; DR N-PSDB; AAA37099.
; XX
; PT New mammalian DNA sequences encoding transmembrane, receptor or
; PT secreted PRO polypeptides, useful for screening of potential peptide or
; PT small molecule inhibitors of the relevant receptor/ligand interactions
; XX
; PS Claim 12; Fig 156; 773pp; English.
; XX
; CC AAA37022 to AAA37144 encode the new isolated human transmembrane,
; CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
; CC transmembrane and receptor PRO proteins can be used for screening of
; CC potential peptide or small molecule inhibitors of the relevant
; CC receptor/ligand interactions. The polypeptides and nucleotide sequences
; CC encoding them have various industrial applications, including uses as
; CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
; CC PCR primers and hybridisation probes used in the isolation of the PRO
; CC polypeptides from the present invention.
; XX
; SQ Sequence 432 AA;
;
; AAY99417 Length: 432 November 25, 2003 13:16 Type: P Check: 6343 ..
aay99417
Query Match 79.1%; Score 1266; DB 1; Length 432;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
;
QY 51 KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWLTAACHCRKHTDVFENKVRAGSD 110
:|||||
Db 199 RVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTAACHCRKHTDVFENKVRAGSD 258
:|||||
QY 111 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
|||||
```

```

Db 259 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 318
QY 171 LWIIGWGFTKQNGKMSDILLOASVQVIDSTRCNADDAAYQGEVTEKMKCAGIPEGGVDT 230
Db 319 LWIIGWGFTKQNGKMSDILLOASVQVIDSTRCNADDAAYQGEVTEKMKCAGIPEGGVDT 378
QY 231 QGDSGGPLMYQSDQWHVVGVISWGYGCGGPSTPGVYTKVSAYLNWIYNVWKAEL 284
Db 379 QGDSGGPLMYQSDQWHVVGVISWGYGCGGPSTPGVYTKVSAYLNWIYNVWKAEL 432

RESULT 2
q9nrs4
; TOIG of: q9nrs4 check: 316 from: 1 to: 437
;
; ID TMS4 HUMAN STANDARD; PRT; 437 AA.
; AC Q9NRS4; Q9NZA5;
; DT 16-OCT-2001 (Rel. 40, Created)
; DT 16-OCT-2001 (Rel. 40, Last sequence update)
; DT 15-SEP-2003 (Rel. 42, Last annotation update)
; DE Transmembrane protease, serine 4 (EC 3.4.21.-) (Membrane-type serine
; DE protease 2) (MT-SP2).
; GN TMPRSS4 OR TMPRSS3.
; OS Homo sapiens (Human).
; OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
; OX NCBI_TaxID=9606;
; RN [1]
; RP SEQUENCE FROM N.A.
; RC TISSUE=Pancreatic carcinoma;
; RX MEDLINE=20283276; PubMed=10825129;
; RA Wallrapp C., Haehnel S., Mueller-Pillasch F., Burghardt B.,
; RA Iwamura T., Ruthenbuenger M., Lerch M.M., Adler G., Gress T.M.;
; RA "A novel transmembrane serine protease (TMPRSS3) overexpressed in
; RA pancreatic cancer.";
; RL Cancer Res. 60:2602-2606 (2000).
; RN [2]
; RP SEQUENCE FROM N.A.
; RA Smeekens S.S., Lorimer D.D., Wang E., Hou J., Linnevers C.;
; RA "MT-SP2, a novel type II membrane serine protease expressed in
; RA trachea, colon, and small intestine: identification, cloning, and
; RA chromosomal localization.";
; RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
; RN [3]
; RP SEQUENCE FROM N.A.
; RC TISSUE=Pancreas;
; RX MEDLINE=22388257; PubMed=12477932;
; RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
; RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
; RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
; RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
; RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
; RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
; RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
; RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
; RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
; RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
; RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
; RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
; RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
; RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
; RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
; RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
; RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
; RT "Generation and initial analysis of more than 15,000 full-length
; RT human and mouse cDNA sequences.";
; RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
; CC -!- FUNCTION: MAY PLAY A ROLE IN TUMOR PROGRESSION, METASTASIS
; CC FORMATION AND TUMOR INVASION.
; CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
; CC -!- TISSUE SPECIFICITY: HIGH LEVELS IN PANCREATIC, GASTRIC, COLORECTAL
; CC AND AMPULLARY CANCER. VERY WEAK EXPRESSION IN NORMAL
; CC GASTROINTESTINAL AND UROGENITAL TRACT.
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - nucleic search, using frame_plus_p2n model
Run on: November 25, 2003, 13:57:34 ; Search time 1.79505 Seconds
(without alignments)
2.016 Million cell updates/sec

Title: US-09-607-745-2
Perfect score: 2342
Sequence: 1 MDPDSQDPLNSLDVKPLRKP.....VYTKVSAYLNWIYNVWKAEL 435

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2 seqs, 4160 residues

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=soft -Q=US09607745.pep -DB=nt.seq -SUFFIX=pto
-OUT=nacompars.res -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1
-END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -NCPU=6 -NO_XLPXY
-NEG_SCORES=0 -LONGLOG -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : nt.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2337	99.8	2081	1 af179224	TOIG of: af179222
2	2324	99.2	2079	1 af216312	TOIG of: af216311
C 3	52.5	2.2	2081	1 af179224	TOIG of: af179222
C 4	46	2.0	2079	1 af216312	TOIG of: af216311

ALIGNMENTS

RESULT 1
af179224
; TOIG of: af179224 check: 328 from: 1 to: 2081
; LOCUS AF179224 2081 bp mRNA linear PRI 08-JUN-2000
; DEFINITION Homo sapiens transmembrane serine protease 3 (TMPRSS3) mRNA, complete cds.
; ACCESSION AF179224
; VERSION AF179224.1 GI:8347148
; KEYWORDS
; SOURCE Homo sapiens (human)
; ORGANISM Homo sapiens
; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2081)
Wallrapp,C., Hahnel,S., Muller-Pillasch,F., Burghardt,B.,
Iwamura,T., Ruthenburger,M., Lerch,M.M., Adler,G. and Gress,T.M.
A novel transmembrane serine protease (TMPRSS3) overexpressed in
pancreatic cancer
Cancer Res. 60 (10), 2602-2606 (2000)
20283276
10825129
2 (bases 1 to 2081)
Wallrapp,C. and Gress,T.M.
Direct Submission
Submitted (20-AUG-1999) Internal Medicine I, University of Ulm,
Robert-Koch-Street 8, Ulm, Baden-Wuerttemberg 89081, Germany
Location/Qualifiers
1. .2081
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q23.3"
/tissue_type="pancreatic carcinoma"
1. .2081
/gene="TMPRSS3"
215. .1528
/gene="TMPRSS3"
/codon_start=1
/product="transmembrane serine protease 3"
/protein_id="AAF74526.1"
/db_xref="GI:8347149"
/translation="MLQDPDSQDPLNSLDVKPLRPRIPMETFRKVGIPILILSLA
SIIIVVLIKVLIDKYFLCGQLHFIIPRKQLDGLDPLGEDEHCVKSFPEGPAV
AVRLSKDRSTLOVLDSATGNWFSACEDNFTALAEACRQMGYSKPTFRAVEIGDDQ
DLDVVEITENSQELMRNNSGCLSGSLVSLHCLACKSLKTPRVVGEEASVDSWFW
QVSIQYDKQHVCGSILDPHWLTAHCFRKHDTDVNWKVRAGSDKLGSPFSLAVAKI
IIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDELTPTATPLWIIIGWFTKQ
NGKMSDILLQASVQVIDTRCNADDAAYQGEVTEKMCAGIPEGGVDTCQGDSCGFLM
YQSDQWHVVGVISWVGCGPSTPGVYTKVSAYLNWIYNVWKAEL"
215. .307
/gene="TMPRSS3"
/note="Region: cytoplasmic domain"
308. .373
/gene="TMPRSS3"
/note="Region: transmembrane domain"
374. .1525
/gene="TMPRSS3"
/note="Region: extracellular domain"
602. .604
/gene="TMPRSS3"
/note="glycosylation site"
746. .748
/gene="TMPRSS3"
/note="glycosylation site"
order(800. .802,1142. .1144)
/gene="TMPRSS3"
/note="disulfide bond"
order(902. .904,950. .952)
/gene="TMPRSS3"
/note="disulfide bond"
947. .949
/gene="TMPRSS3"
/note="active site"
1082. .1084
/gene="TMPRSS3"
/note="active site"
order(1280. .1282,1328. .1330)
/gene="TMPRSS3"
/note="disulfide bond"
1373. .1375
/gene="TMPRSS3"
/note="active site"
BASE COUNT 484 a 597 c 576 g 424 t
ORIGIN

AF179224 Length: 2081 November 25, 2003 13:15 Type: N Check: 328 ..
af179224

Alignment Scores:

Pred. No.: 0 Length: 2081
Score: 2337.00 Matches: 434
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 1 Gaps: 0

US-09-607-745-2 (1-435) x af179224 (1-2081)

```

QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
DB 224 GATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAACCCCTGCGCAAAACCCCGT 283
QY 22 IleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSerLeu 41
DB 284 ATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGCCTG 343
QY 42 AlaSerIleIleValValValLeuIleValIleValIleLeuAspLysTyrPheLeu 61
DB 344 GCGAGTATCATATTGTTGCTTCATCAAGGTGATTCTGGATAATACTACTTCTTC 403
QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
DB 404 TCGGGGAGCCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGGAGAGCTGGACTGT 463
QY 82 ProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaValAla 101
DB 464 CCTTGGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTGCGAGTGGCA 523
QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
DB 524 GTCCGCCTCTCCAAGGACCGATCCACACTGCAGGTGTGGACTTCGGCCACAGGGAACTGG 583
QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
DB 584 TTCTCTGCCTGTTTCGACAACTTCACAGAAAGCTTCGCTGAGACAGCCTGTAGGCAATG 643
QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
DB 644 GGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCAGGATCTGGAT 703
QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
DB 704 GTTGTGTAATCACAGAAAACAGCAGGAGCTTCGATCGGAACCTCAAGTGGGCCCTGT 763
QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
DB 764 CTCTCAGGCTCCCTGCTCCCTGCACTGTCTTGCCTGTGGAGAGCCTGAAGACCCCC 823
QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
DB 824 CGTGTGGTGGTGGGAGGAGGCTCTGTGGAATCTTGGCCTTGGCAGGTTCAGCATCCAG 883
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
DB 884 TACGACAAACAGCACGCTCTGTGGAGGAGCATCTTCGACCCCTTCCTGCTCCTCAGCGCA 943
QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
DB 944 GCGCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGCGAGGCTCAGAC 1003
QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
DB 1004 AAACCTGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTAATCAATCAACCCC 1063
QY 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
DB 1064 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCTCCACTCACTTTCTCA 1123

```

```

QY 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
DB 1124 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1183
QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
DB 1184 CTCTGGATCATTTGGATGGGCTTTACGAGCAGAAATGAGGGAAGATGTCTGACATACTG 1243
QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 361
DB 1244 CTGACGGCTCAGTCCAGTCCATGACAGCACACGGTCAATGCAGACGATGCGTACCAG 1303
QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
DB 1304 GGGGAAGTCACCGAGAAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGACACCTGC 1363
QY 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
DB 1364 CAGGTGACAGTGGTGGGCCCTCTGATACCAATCTGACAGTGGCATGTGGTGGGCATC 1423
QY 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
DB 1424 GTTAGCTGGGCTATGGCTGCGGGGCCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1483
QY 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
DB 1484 GCCTATCTCACTGATCTACAATGCTGGAAGGTGAGCTG 1525

```

RESULT 2

af216312

TOIG of: af216312 check: 3320 from: 1 to: 2079

```

; LOCUS AF216312 2079 bp mRNA linear PRI 07-FEB-2000
; DEFINITION Homo sapiens type II membrane serine protease mRNA, complete cds.
; ACCESSION AF216312
; VERSION AF216312.1 GI:6911218
; KEYWORDS
; SOURCE Homo sapiens (human)
; ORGANISM Homo sapiens
; MAMMALIA; EUTHERIA; PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
; REFERENCE 1 (bases 1 to 2079)
; AUTHORS Smeekens, S.S., Lorimer, D.D., Wang, E., Hou, J. and Linnevers, C.
; TITLE MT-SP2, a novel type II membrane serine protease expressed in
; trachea, colon, and small intestine: identification, cloning, and
; chromosomal localization
; JOURNAL Unpublished
; REFERENCE 2 (bases 1 to 2079)
; AUTHORS Smeekens, S.S., Lorimer, D.D., Wang, E., Hou, J. and Linnevers, C.
; TITLE Direct Submission
; JOURNAL Submitted (14-DEC-1999) Axys Pharmaceuticals, Inc, 180 Kimball Way,
; South San Francisco, CA 94080, USA
; FEATURES
; Location/Qualifiers
; 1..2079
; /organism="Homo sapiens"
; /mol_type="mRNA"
; /db_xref="taxon:9606"
; 251..1522
; /note="MT-SP2"
; /codon_start=1
; /product="type II membrane serine protease"
; /protein_id="AAF31436.1"
; /db_xref="GI:6911219"
; /translation="MSNPANPVPWPSPESVGIPIIALLSIASIIIVVVLKVLID
; KYFLCGQPLHFIPRKQCDGELDCPLGEDEHCVKSPFEGPAVAVRLSKDRSTLQVL
; DSATGNWFSACFDNFTAEALAEACRQMGYSKPTFAVEIGPDQDLVVEITENSQEL
; RMNRSSGCLSGSLHCLACGSKLTPRVVVGGEASVDSWPWQVSIQYDKQHVCGG
; SILDPHWLTAHCFRKHDTFVNWKVRAGSKLGSFSLAVAKIIIEFNPMYPKDND
; IALMKLQPLTFSGTVPICLPFFDEELTPATLWIGFTKONGGKMSDILLQASV
; QVIDSTRCNADDAQQEVEKMKCAGIPEGGVDTCCQSDSGGPLMYQSDQWVVGVSW
; GYCGGSPSTPGVYTKVSAYLWNIYNVWKAEL"
; BASE COUNT 489 a 594 c 575 g 421 t

```


RESULT 3

```

/translation="MLQPDSDQPLNSLDVPLKPRIPMETFRKVGIPILIALLSLA
SIIIVVLIKVLKDYFLCGQPLHFIIPKQLCDGELDCPLGEDEHCVKSFPEGPAY
AVRLSKDRSTLQVLDSATGNWFSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQ
DLDVVEITENSQELMRNMSGPCLSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPW
QVSIQYDKQHVCGGSLDPHWLTAACHCFKHTDVFNWVKVRAGSKLGSFPFSLAVAKI
IIIEFNPMYPKNDIALMKLOFPLTFSGTVRPICLPFFDEELTPATPLWIIIGWFTTKQ
NGGKMSDILLQASVOVIDSTRCNADDAAYQGEVTEKMMCAIGIPEGGVDTCCQDSSGGPLM
YQSDQWHVVGIVSWGCGGPSTPGVYTKVSAYLNIYNVWKAEL"
215. .307
/gene="TMPRSS3"
/region="Region: cytoplasmic domain"
308. .373
/gene="TMPRSS3"
/region="Region: transmembrane domain"
374. .1525
/gene="TMPRSS3"
/region="Region: extracellular domain"
602. .604
/gene="TMPRSS3"
/region="glycosylation site"
746. .748
/gene="TMPRSS3"
/region="glycosylation site"
order(800. .802,1142. .1144)
/gene="TMPRSS3"
/region="disulfide bond"
order(902. .904,950. .952)
/gene="TMPRSS3"
/region="disulfide bond"
947. .949
/gene="TMPRSS3"
/region="active site"
1082. .1084
/gene="TMPRSS3"
/region="active site"
order(1280. .1282,1328. .1330)
/gene="TMPRSS3"
/region="disulfide bond"
1373. .1375
/gene="TMPRSS3"
/region="active site"
484 a 597 c 576 g 424 t
; BASE COUNT
; ORIGIN
; AF179224 Length: 2081 November 25, 2003 13:15 Type: N Check: 328 ..
af179224
Alignment Scores:
Pred. No.: 0 Length: 2081
Score: 52.50 Matches: 52
Percent Similarity: 35.62% Conservative: 26
Best Local Similarity: 23.74% Mismatches: 86
Query Match: 2.24% Indels: 57
DB: 1 Gaps: 10
US-09-607-745-2 (1-435) x af179224 (1-2081)
QY 23 ProMetGluThrPheArgLysValGlyIle-----ProIleIleIleAlaLeuLeu 39
Db 684 CCAATCTCCACAGCTCTGAAAGTGGGTTTGCTGTGTAGCCCATCTGCCTACAGGCTGTC 625
QY 40 SerLeuAlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyr--- 58
Db 624 TCAGCGAGAGCT-----TCTGTGAAGTTGTGGAACAGGCA 589
QY 59 -----TyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAsp 76
Db 588 GAGAACCAAGTTCCC-TGTGGCCGAGTCCAGCACCTGCAGTGTGGATCGGTCTT----- 536
QY 77 GlyGluLeuAspCys-----ProLeuGlyGluAspGluGluHisCysValLysSer 93
Db 535 GGAGAGCGGCGACTGCCACTGCAGGCCCTTCGGGGAAGCTCTTGACACAGTGCTCCTCGTC 476
```

```

QY 94 PheProGluGlyProAla---ValAlaValArgLeuSerLysAspArgSerThrLeuGln 112
Db 475 CTCCTCCCAAGGACAGTCCAGCTCTCCGTCACACAGCTGCTTCTCCGAGTGAAGTGAG 416
QY 113 ValLeu-----AspSerAlaThr 118
Db 415 AGGCTGCCCGCAGAGGAGTAGTATTATCCAGAATCACCTTGATGAGGACCAACCAAT 356
QY 119 GlyAsnTrpPheSerAla-----CysPheAspAsnPheThrGluAlaLeuAlaGluThr 136
Db 355 GATGATACTCGCCAGGCTCAGTAGTGTCTATGATGATGGGATCCCCACCTTCTGAAGGT 296
QY 137 AlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyPro 156
Db 295 CTCCATGGGGATACGGGGTTGGCGAGGGGTTGACATCGAGGCTGTT----- 248
QY 157 AspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsn 176
Db 247 ---CAGAGGTTGATCAGTGTCTCAGGATCCTGTAAACAT----- 215
QY 177 SerSerGlyProCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLys 196
Db 214 -----GCTGGCTCTGTGATCTCCTCCGCTCCTCCCTAGCCTGGAGCT 173
QY 197 SerLeuLysThrProArgValValGlyGlyGluGluAlaSerValAspSerTyrPro 215
Db 172 GAGATTGTACCCCAAGGCA-GCAGGAGGGGCTCCACACAGGTCCTGGCTGGCCA 117
RESULT 4
af216312/c
; TOIG of: af216312 check: 3320 from: 1 to: 2079
; LOCUS AF216312 2079 bp mRNA linear PRI 07-FEB-2000
; DEFINITION Homo sapiens type II membrane serine protease mRNA, complete cds.
; ACCESSION AF216312
; VERSION AF216312.1 GI:6911218
; KEYWORDS
; SOURCE Homo sapiens (human)
; ORGANISM Homo sapiens
; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
; REFERENCE 1 (bases 1 to 2079)
; AUTHORS Smeekens,S.S., Lorimer,D.D., Wang,E., Hou,J. and Linnevers,C.
; TITLE MT-SP2, a novel type II membrane serine protease expressed in
; trachea, colon, and small intestine: identification, cloning, and
; chromosomal localization
; JOURNAL Unpublished
; REFERENCE 2 (bases 1 to 2079)
; AUTHORS Smeekens,S.S., Lorimer,D.D., Wang,E., Hou,J. and Linnevers,C.
; TITLE Direct Submission
; JOURNAL Submitted (14-DEC-1999) Axy's Pharmaceuticals, Inc, 180 Kimball Way,
; South San Francisco, CA 94080, USA
; FEATURES
; Location/Qualifiers
; 1..2079
; /organism="Homo sapiens"
; /mol_type="mRNA"
; /db_xref="taxon:9606"
; 251..1522
; /note="MT-SP2"
; /codon_start=1
; /product="type II membrane serine protease"
; /protein_id="AAF31436.1"
; /db_xref="GI:6911219"
; /translation="MSNPANVSPWRPSESVGIPILIALLSLSTIIIVVLIKVLID
KYFFLCGQPLHFIIPKQLCDGELDCPLGEDEHCVKSFPEGPAYAVRLSKDRSTLQVL
DSATGNWFSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDLVDVVEITENSQEL
MRNMSGPCLSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPVQVSIQYDKQHVCGG
SILDPHWLTAACHCFKHTDVFNWVKVRAGSKLGSFPFSLAVAKIIEFNPMYPKNDND
IALMKLOFPLTFSGTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASV
QVIDSTRCNADDAAYQGEVTEKMMCAIGIPEGGVDTCCQDSSGGPLMQSDQWHVVGIVS
W"
; BASE COUNT 489 a 594 c 575 g 421 t
; ORIGIN
```

/ ; AF216312 Length: 2079 November 25, 2003 13:16 Type: N Check: 3320 ..
af216312

Alignment Scores:

Pred. No.: 0 Length: 2079
Score: 46.00 Matches: 25
Percent Similarity: 33.59% Conservative: 18
Best Local Similarity: 19.53% Mismatches: 47
Query Match: 1.96% Indels: 38
DB: 1 Gaps: 4

US-09-607-745-2 (1-435) x af216312 (1-2079)

QY 129 PheThrGluAlaLeuAlaGluThrAlaCysArgGln-----MetGlyTyrSerSerLys 146
Db 955 TTTCCTGAAGCAGTG-----GGCTGCCCTGAGGACCCAGTGGGGTCCAGGATGCT 905
QY 147 ProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAspValGluIleThr 166
Db 904 CCTCCACAGACGCTGCTGTTGTCGTACTGGATGCTGACCTG----- 863
QY 167 GluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCysLeuSerGlySerLeu 186
Db 862 -----CCAGGCCAAGATCCACAGAGGCCTC 836
QY 187 ValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrProArgValValGlyGly 206
Db 835 CTCCCCACCCACACAGCGGGGCTCTTCAGGCTCTTCCACAGGCAAGACAGTGCAGGGA 776
QY 207 GluGluAlaSerValAspSerTrpProTrpGlnValSerIleGlnTyrAspLysGlnHis 226
Db 775 GACCAGGGAGCCTGAGACAGAGGCCCACTTGAGTT----- 740
QY 227 ValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAlaAlaHisCysPheArg 246
Db 739 -----CCGATGCGAAGCTCCTGGCTGTTTCTGTGATTTC 704
QY 247 LysHisThrAspValPheAsnTrp 254
Db 703 AACACATCCAGATCCTGCTGG 680

Search completed: November 25, 2003, 13:57:41
Job time : 5.79505 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 25, 2003, 13:57:34 ; Search time 1.20495 Seconds
(without alignments)
2.016 Million cell updates/sec

Title: US-09-607-745-9
Perfect score: 1600
Sequence: 1 MDSKGSSQKSRLLLLLVSN.....LNWIYVWKAELSRHHHHH 292

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2 seqs, 4160 residues

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=soft -Q=US09607745.pep -DB=nt.seq -SUFFIX=pto
-OUT=nacompare.res -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1
-END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -NCEU=6 -NO XLPXY
-NEG SCORES=0 -LONGLOG -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : nt.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1266	79.1	2079	1	af216312	TOIG of: af21631
2	1266	79.1	2081	1	af179224	TOIG of: af17922
c 3	37.5	2.3	2079	1	af216312	TOIG of: af21631
c 4	37.5	2.3	2081	1	af179224	TOIG of: af17922

ALIGNMENTS

RESULT 1
af216312

; TOIG of: af216312 check: 3320 from: 1 to: 2079

; LOCUS AF216312 2079 bp mRNA linear PRI 07-FEB-2000
; DEFINITION Homo sapiens type II membrane serine protease mRNA, complete cds.
; ACCESSION AF216312
; VERSION AF216312.1 GI:6911218

; KEYWORDS
; SOURCE Homo sapiens (human)

; ORGANISM Homo sapiens

; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

; REFERENCE 1 (bases 1 to 2079)

; AUTHORS Smeekens,S.S., Lorimer,D.D., Wang,E., Hou,J. and Linnevers,C.

; TITLE MT-SP2, a novel type II membrane serine protease expressed in
; trachea, colon, and small intestine: identification, cloning, and
; chromosomal localization

; JOURNAL Unpublished

; REFERENCE 2 (bases 1 to 2079)

; AUTHORS Smeekens,S.S., Lorimer,D.D., Wang,E., Hou,J. and Linnevers,C.

; TITLE Direct Submission

; JOURNAL Submitted (14-DEC-1999) Axys Pharmaceuticals, Inc, 180 Kimball Way,
; South San Francisco, CA 94080, USA

; FEATURES Location/Qualifiers

; source

; /organism="Homo sapiens"

; /mol_type="mRNA"

; /db_xref="taxon:9606"

; 251..1522

; /note="MT-SP2"

; /codon_start=1

; /product="type II membrane serine protease"

; /protein_id="AAF31436.1"

; /db_xref="GI:6911219"

; /translation="MSNPANPVSVPWPSVGIPIIIALLSLASIIIVVLIKVIDL
; KYFLCGQPLHPIPRKQLCDGELDCPLGEDEHCVKSPFEGPAVRLSKDRSTLQVL

; DSATGNWFSACFDNTEALAEATACROMGYSSKPTFRAVEIGPDQDLDDVVEITENSQEL

; RMRNSSGPCLSGLVSLHCLACGKSLKTRVVGEEASVDSWPWQVSIQYDKQHVCGG

; SILDEHWLTAHCFRKHTDVENWVKVAGSKLGSFSLAVAKIIIEFNPMYPKND

; IALMKLQPLTFSGTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLQASV


```

; misc_feature      /note="disulfide bond"
;                  1373..1375
;                  /gene="TMPRSS3"
;                  /note="active site"
;
; BASE COUNT      484 a 597 c 576 g 424 t
; ORIGIN
;
; AF179224 Length: 2081 November 25, 2003 13:15 Type: N Check: 328 ..
af179224
```

```

Alignment Scores:
Pred. No.:      0      Length:      2081
Score:          1266.00  Matches:      228
Percent Similarity: 98.29%  Conservative: 2
Best Local Similarity: 97.44%  Mismatches: 4
Query Match:      79.12%  Indels: 0
DB:               1      Gaps: 0
```

US-09-607-745-9 (1-292) x af179224 (1-2081)

```

QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
Db 824 CGTGTGGTGGTGGGAGGAGGCCTCTGTGATTTCTTGGCCTTGGCAGTCCAGTCCAG 883
QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
Db 884 TAGACAAACAGCACGCTCTGTGGAGGGAGCATCTGGACCCCACTGGTCTCTACGGCA 943
QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
Db 944 GCCCACTGCTTCAGGAAACATACCGATGTGTCAACTGGAAGGTGCGGCGAGGCTCAGAC 1003
QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
Db 1004 AAACCTGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATGAATTCAACCC 1063
QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
Db 1064 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCGCACTCATTCTCA 1123
QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
Db 1124 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1183
QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
Db 1184 CTCTGGATCATTTGGATGGGGCTTTACGAGCAGAGATGGAGGAGAGATGTCTGACATACTG 1243
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
Db 1244 CTGCAGGGCTCAGTCCAGGTCATTGACAGCACACCGTGCATATGCAGACGATGCGTACCAG 1303
QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
Db 1304 GGGGAAGTCACCGAAGATGATGTGTGAGGCATCCCGGAGGGGTGTGGACACCTGC 1363
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
Db 1364 CAGGGTGACAGTGGTGGGCCCTGTATGATACCAATCTGACCAAGTGGCATGTGTGGGCATC 1423
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
Db 1424 GTTAGCTGGGGCTATGGCTGGGGGCCGAGCACCCAGAGATATACACCAAGGTCTCA 1483
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1484 GCCTATCTCACTGATCTACAATGTCTGGAAGGCTGAGCTG 1525
```

RESULT 3

af216312/c

; TOIG of: af216312 check: 3320 from: 1 to: 2079

, LOCUS AF216312 2079 bp mRNA linear PRI 07-FEB-2000

```

; DEFINITION Homo sapiens type II membrane serine protease mRNA, complete cds.
; ACCESSION AF216312
; VERSION AF216312.1 GI:6911218
; KEYWORDS
; SOURCE Homo sapiens (human)
; ORGANISM Homo sapiens
; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
; REFERENCE 1 (bases 1 to 2079)
; AUTHORS Smeekens,S.S., Lorimer,D.D., Wang,E., Hou,J. and Linnevers,C.
; TITLE MT-SP2, a novel type II membrane serine protease expressed in
; trachea, colon, and small intestine: identification, cloning, and
; chromosomal localization
; JOURNAL Unpublished
; REFERENCE 2 (bases 1 to 2079)
; AUTHORS Smeekens,S.S., Lorimer,D.D., Wang,E., Hou,J. and Linnevers,C.
; TITLE Direct Submission
; JOURNAL Submitted (14-DEC-1999) Axys Pharmaceuticals, Inc, 180 Kimball Way,
; South San Francisco, CA 94080, USA
; FEATURES
; Location/Qualifiers
; 1..2079
; /organism="Homo sapiens"
; /mol_type="mRNA"
; /db_xref="taxon:9606"
; 251..1522
; /note="MT-SP2"
; /codon_start=1
; /product="type II membrane serine protease"
; /protein_id="AAF31436.1"
; /db_xref="GI:6911219"
; /translation="MSNP CANPVS PWRPSES VGPIIIALLSLASIIIVVLIKVILD
; KYFLCGQLHFIPRKQLCDGELDCPLGDEDEHCVKSFPEGPAVAVRLSKDRSTLQVL
; DSATGNWFSACFDNFTALAEATACROMGYSSKPTFAVEIGPDQDLDVVEITENSQEL
; RMRNSSGCLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPWQVSIQYDKQHVCGG
; SILDPHWLTAACHCRKHTDVFNWKVRAGSKLGSFSLAVAKIIIEFNPMYPKDND
; IALMKLQFPLTFSGTVRPICLPFFDELTPTATPLMIIGWFTKONGGKMSDILLQASV
; QVIDSTRCNADDAAYQGEVTEKMCAGIPEGGVDTCCQSDSGGPLMYQSDQWHVVGVISW
; GYGGGSPSTPGVTVTKVSAYLNWIYNVWKAEI"
; BASE COUNT 489 a 594 c 575 g 421 t
; ORIGIN
;
; AF216312 Length: 2079 November 25, 2003 13:16 Type: N Check: 3320 ..
af216312
```

```

Alignment Scores:
Pred. No.:      0      Length:      2079
Score:          37.50  Matches:      34
Percent Similarity: 35.66%  Conservative: 17
Best Local Similarity: 23.78%  Mismatches: 49
Query Match:      2.34%  Indels: 43
DB:               1      Gaps: 10
```

US-09-607-745-9 (1-292) x af216312 (1-2079)

```

QY 150 SerGlyThrValArgProIleCysLeuProPhePheAspGlu----- 163
Db 1864 TCTGGGCTTTTACAGACAGCAGCTGCTTCCATTCACTAGTGTGGAAAGTTCTTCTTGGC 1805
QY 164 -----GluLeuThrProAlaThrProLeuTrpIleIleGlyTrp-----GlyPhe 178
Db 1804 TTAGCAATACCCCTGAGACCTTGTTCAGTGGGTGTGTCTCTCCCTGGGATGTGGGAGC 1745
QY 179 ThrLysGlnAsnGlyGlyLysMetSerAspIleLeu----- 190
Db 1744 ACCAAGTGTGGCCGAGCTAGGCTGTGACTTCTCTGGGCGCCTCTGGGCTCGAGGGT 1685
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
Db 1684 CTCATTACAGGAATTGAGGCCCTTTGTCTGTCTCCAGA---AATGCTGAGGCTGTGGGCAGA 1628
QY 211 GlyGluValThrGlu---LysMetMetCysAla-----GlyIlePro----- 223
Db 1627 GGGGTGTACCAAGGGGACTCTTGTCTGTCTGCTACTTGGGGGATCCCCAGGTGGGCA 1568
```



```
Qy 224 -----GluGly---GlyValAspThrCysGlnGly----- 232
Db 1567 GGCAGGAGGAGCGGCTCCAGCACTGCAAGGGGAGCAGCATTACAGCTCAGCCTT 1508
Qy 233 ---AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValvalGlyIleVal 251
Db 1507 CCAGACATTGTAGATCCAGTTGATAGATAGCTGAGAC-----CTTGTGTATAC 1460
Qy 252 SerTrpGly 254
Db 1459 TCCTGGGT 1451

RESULT 4
af179224/c
; TOIG of: af179224 check: 328 from: 1 to: 2081
;
; LOCUS AF179224 2081 bp mRNA linear PRI 08-JUN-2000
; DEFINITION Homo sapiens transmembrane serine protease 3 (TMPRSS3) mRNA,
; complete cds.
; ACCESSION AF179224
; VERSION AF179224.1 GI:8347148
; KEYWORDS
; SOURCE
; ORGANISM Homo sapiens (human)
; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
; REFERENCE 1 (bases 1 to 2081)
; AUTHORS Wallrapp,C., Hahnel,S., Muller-Pillasch,F., Burghardt,B.,
; Iwamura,T., Ruthenburger,M., Lerch,M.M., Adler,G. and Gress,T.M.
; TITLE A novel transmembrane serine protease (TMPRSS3) overexpressed in
; pancreatic cancer
; JOURNAL Cancer Res. 60 (10), 2602-2606 (2000)
; MEDLINE 20283276
; PUBMED 10825129
; REFERENCE 2 (bases 1 to 2081)
; AUTHORS Wallrapp,C. and Gress,T.M.
; TITLE Direct Submission
; JOURNAL Submitted (20-AUG-1999) Internal Medicine I, University of Ulm,
; Robert-Koch-Street 8, Ulm, Baden-Wuerttemberg 89081, Germany
; FEATURES
; source
; location/Qualifiers
; 1..2081
; /organism="Homo sapiens"
; /mol_type="mRNA"
; /db_xref="taxon:9606"
; /chromosome="11"
; /map="11q23.3"
; /tissue type="pancreatic carcinoma"
; 1..2081
; /gene="TMPRSS3"
; 215..1528
; /gene="TMPRSS3"
; /codon_start=1
; /product="transmembrane serine protease 3"
; /protein_id="AAF74526.1"
; /db_xref="GI:8347149"
; /translation="MLQDPDSQPLNSLDVVKPLRPKPRIPMETFRKVGIPPIIIALLSLA
; SIIIVVLIKVIDKYFLCGQLHPIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAV
; AVRLSKDSTLQVLDSATGNWFSACFDNFTALAEATACRQMGYSKPTFRVAVIGPDQ
; DLDVVEITENSQELMRNSSGPCLSGSLVSLHCLACGSLKTPRVVGEEASVDSWPW
; QVSIQDKVHVGGSILDPHWVLTAAHCFRKHDTDFNWKVRAGSDKLGSPFSLAVAKI
; IIEFPMPKDNIDIALMKLQPLTFSGTVRPICLPFDEELTPATPLWIIGWGTQK
; NGKMSDILLQASVQVIDSTRCNADDAVQGEVTEKMKMCAIPEGGVDTCCQSDSGGPLM
; YQSDQHVTVGLVSWVGCGGPGSTPGVYTKVSAYLNIYNWKAEL"
; 215..307
; /gene="TMPRSS3"
; /note="Region: cytoplasmic domain"
; 308..373
; /gene="TMPRSS3"
; /note="Region: transmembrane domain"
; 374..1525
; /gene="TMPRSS3"
; /note="Region: extracellular domain"
```

```
misc_feature 602..604
; /gene="TMPRSS3"
; /note="glycosylation site"
misc_feature 746..748
; /gene="TMPRSS3"
; /note="glycosylation site"
misc_feature order(800..802,1142..1144)
; /gene="TMPRSS3"
; /note="disulfide bond"
misc_feature order(902..904,950..952)
; /gene="TMPRSS3"
; /note="disulfide bond"
misc_feature 947..949
; /gene="TMPRSS3"
; /note="active site"
misc_feature 1082..1084
; /gene="TMPRSS3"
; /note="active site"
misc_feature order(1280..1282,1328..1330)
; /gene="TMPRSS3"
; /note="disulfide bond"
misc_feature 1373..1375
; /gene="TMPRSS3"
; /note="active site"
; BASE COUNT 484 a 597 c 576 g 424 t
; ORIGIN
;
; AF179224 Length: 2081 November 25, 2003 13:15 Type: N Check: 328
af179224
Alignment Scores:
Pred. No.: 0 Length: 2081
Score: 37.50 Matches: 36
Percent Similarity: 33.33% Conservative: 14
Best Local Similarity: 24.00% Mismatches: 55
Query Match: 2.34% Indels: 46
DB: 1 Gaps: 8

US-09-607-745-9 (1-292) x af179224 (1-2081)
Qy 138 IleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCys 157
Db 1952 CTTGCTCTAGAGCTTGGGATGGGTGAAACGACAGGGCTGGCGCAGACCC----- 1899
Qy 158 LeuProPheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGly 177
Db 1898 -----TTTCCTTCTCCTCTCCAGCCACAC-----GTGATCTGGGCT 1863
Qy 178 PheThrLysGlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGln--- 196
Db 1862 TTACAAGACAG-----CCTGCTTCCATTCCAGTAG 1833
Qy 197 -----ValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGlnGlyGluValThr 214
Db 1832 TGTGGAAAGTCTCTTGGCTTAGCAATACCCCTGAGACCTTGTTCAGTGGGCTGTGT 1773
Qy 215 GluLysMetMetCysAlaGlyIlePro-----GluGlyGlyValAspThrCysGln 231
Db 1772 CTCTCCCTGGGATGCTG-GGAGCACCAGTGTGGCCGAGTAGGGCTGTGACTTCTCTCT 1714
Qy 232 Gly-----AspSerGlyProLeuMetTyrGlnSer 242
Db 1713 GGGCGCTCTGGGCTGCGAGGGTCTTTATAGGAATTAGGCCCTTTGTCTGCTCAAGAA 1654
Qy 243 AspGlnTrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThr 262
Db 1653 -----ATGCTGAGGCTGTGGCAGAGGGGTGT-----ACC 1624
Qy 263 ProGlyValThrLysValSerAlaTyr 272
Db 1623 CAAGGGGACTTGTCTGTGTCTGACTTT 1594
```

Tue Nov 25 13:58:05 2003

nacompare.res

Page 9

Search completed: November 25, 2003, 13:57:43
Job time : 3.20495 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

QM protein - protein search, using sw model

Run on: November 23, 2003, 07:38:25 ; Search time 64.0234 Seconds
(without alignments)
1078.451 Million cell updates/sec

Title: US-09-607-745-2

Perfect score: 2342

Sequence: 1 MDPDSQPLNSLDVKPLRKP.....VYTKVSAYLNWIYVWKAEI 435

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2342	100.0	435	22 AAY72558	Human seripancrin
2	2342	100.0	435	23 AAG78577	Human serine prote
3	2342	100.0	435	24 ABU04912	Human expressed pr
4	2342	100.0	435	24 ABU04931	Human expressed pr
5	2338	99.8	435	20 AAY06437	Human protease HUP
6	2338	99.8	435	24 ABU04930	Human expressed pr
7	2337	99.8	437	23 AAG79359	CJA8 preferred seq
8	2337	99.8	437	24 ABU04915	Human expressed pr
9	2337	99.8	437	24 ABU04920	Human expressed pr

10	2337	99.8	437	24	ABU04935	Human expressed pr
11	2337	99.8	437	24	ABU04936	Human expressed pr
12	2329	99.4	492	22	AAY72559	Human seripancrin
13	2329	99.4	492	24	ABU04913	Human expressed pr
14	2329	99.4	492	24	ABU04932	Human expressed pr
15	2319	99.0	437	22	AAE06931	Human membrane-tyr
16	2319	99.0	437	24	ABU04917	Human expressed pr
17	2319	99.0	437	24	ABU04922	Human expressed pr
18	2297.5	98.1	432	21	AAU99417	Human PRO1570 (UNQ
19	2297.5	98.1	432	22	AAU29188	Human PRO polypept
20	2297.5	98.1	432	22	AAE87581	Human PRO1570. Ho
21	2297.5	98.1	432	22	AAE66166	Protein of the inv
22	2297.5	98.1	432	23	ABG95906	Human secreted/tra
23	2297.5	98.1	432	23	AAU76535	Tumour-associated
24	2297.5	98.1	432	24	ABU71276	Human PRO1570 prot
25	2297.5	98.1	432	24	ABU71561	Human secreted pol
26	2297.5	98.1	432	24	ABU72007	Novel human secret
27	2297.5	98.1	432	24	ABU72164	Human PRO polypept
28	2297.5	98.1	432	24	ABU65733	Human secreted/tra
29	2297.5	98.1	432	24	ABU66066	Novel human secret
30	2297.5	98.1	432	24	ABU67570	Human secreted/tra
31	2297.5	98.1	432	24	ABU65428	Human PRO polypept
32	2297.5	98.1	432	24	ABU58564	Human PRO polypept
33	2297.5	98.1	432	24	ABU56100	Human secreted/tra
34	2297.5	98.1	432	24	ABU57095	Human PRO polypept
35	2297.5	98.1	432	24	ABU04921	Human expressed pr
36	2297.5	98.1	432	24	ABU04929	Human expressed pr
37	2297.5	98.1	432	24	ABU04934	Human expressed pr
38	2297.5	98.1	432	24	ABU10674	Human secreted/tra
39	2287	97.7	428	23	ABG96430	Human ovarian canc
40	2251	96.1	521	22	AAU16975	Human novel secret
41	2251	96.1	521	24	ABU04925	Human expressed pr
42	2242	95.7	420	24	ABU56625	Lung cancer-associ
43	2198	93.9	423	22	AAE06944	Human transmembran
44	2198	93.9	423	24	ABU04918	Human expressed pr
45	2198	93.9	423	24	ABU04919	Human expressed pr

ALIGNMENTS

RESULT 1

AAU72558
ID AAY72558 standard; Protein; 435 AA.

XX AAY72558;

XX AAY72558;

DT 02-MAY-2001 (first entry)

XX Human seripancrin protein.

KW Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;
KW arthritis; chronic obstructive pulmonary disorder; COPD; cancer;
KW osteoporosis; aberrant wound healing; angiogenesis; diabetes;
KW inflammatory disorder; stroke; cerebroprotective; vulnary; osteopathic.

XX Homo sapiens.

OS WO200104141-A2.

PN 18-JAN-2001.

XX (04-JUL-2000; 2000WO-EP06211.)

PR 12-JUL-1999; 99EP-0113428.

XX (MERE) MERCK PATENT GMBH.

PI Suendermann B, Hofmann U, Matzku S, Wilbert O;

XX WPI; 2001-147177/15.

DR N-PSDB; AAD02556.

XX New extracellular serine protease Seripancrin, useful for treating
PT cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease,
PT diabetes, inflammatory disorders, stroke, angiogenesis and aberrant
PT wound healing
XX
PS Claim 2; Page 39-40; 45pp; English.
XX
CC The present invention relates to seripancrin polynucleotides, and
CC polypeptides encoded by them. Seripancrin are members of serine
CC protease family. This protein contains a transmembrane domain,
CC a low density lipoprotein (LDL) domain, protease domain and a
CC scavenger receptor cysteine-rich (SRCR) domain. The LDL and SRCR
CC domains help to define the specificity of seripancrin's intra and
CC intermolecular interactions. The polynucleotides and polypeptides
CC of the invention are useful for treating and diagnosing diseases
CC such as arthritis, chronic obstructive pulmonary disorder (COPD),
CC cancer, osteoporosis, aberrant wound healing, angiogenesis,
CC inflammatory disorders, diabetes, stroke and cardiovascular diseases.
CC Seripancrin genes are useful in chromosome localisation studies,
CC as tools for tissue expression studies and also in gene therapy.
CC The polypeptides of the invention are used for identifying agonists
CC and antagonists useful for treating conditions associated with
CC seripancrin imbalance. These polypeptides are also useful as vaccines.
CC The present sequence is seripancrin protein. The
CC seripancrin gene is located on human chromosome 11q22-q23.
XX
SQ Sequence 435 AA;

Query Match 100.0%; Score 2342; DB 22; Length 435;
Best Local Similarity 100.0%; Pred. No. 3.3e-194;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVKPLRKPRIPIPMETFRKVGIPITIALLSASIIIVVVLKVIDKYFF 60
Db |||||||
QY 1 MDPDSQPLNSLDVKPLRKPRIPIPMETFRKVGIPITIALLSASIIIVVVLKVIDKYFF 60
Db |||||||
QY 61 LCGQLHFIPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGN 120
Db |||||||
QY 61 LCGQLHFIPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGN 120
Db |||||||
QY 121 WFSACFDNFTALAEATACRQMGYSKPTFRAVEIGDQDLVVEITENSQELMRNSSGP 180
Db |||||||
QY 121 WFSACFDNFTALAEATACRQMGYSKPTFRAVEIGDQDLVVEITENSQELMRNSSGP 180
Db |||||||
QY 181 CLSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGSILDPHWLVT 240
Db |||||||
QY 181 CLSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGSILDPHWLVT 240
Db |||||||
QY 241 AAHCFRKHTDVFNKVRAGSKLGSFPPLAVAKIIIEFNPMYPKDNIALMKLQFPLTF 300
Db |||||||
QY 241 AAHCFRKHTDVFNKVRAGSKLGSFPPLAVAKIIIEFNPMYPKDNIALMKLQFPLTF 300
Db |||||||
QY 301 SGTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDDAY 360
Db |||||||
QY 301 SGTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDDAY 360
Db |||||||
QY 361 QGEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQWHVVGIVSWGCGGPSTPGVYTKV 420
Db |||||||
QY 361 QGEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQWHVVGIVSWGCGGPSTPGVYTKV 420
Db |||||||
QY 421 SAYLNWIYNVWKAEL 435
Db |||||||
QY 421 SAYLNWIYNVWKAEL 435
Db |||||||

RESULT 2
AAG78577
ID AAG78577 standard; Protein; 435 AA.

XX AAG78577;

AC AAG78577;
XX 07-MAY-2002 (first entry)

XX Human serine protease D-G amino acid sequence.
DE
XX Serine protease; D-G; human; zymogen; enzyme; cytostatic;
KW antiinflammatory; dermatological; anticoagulation; cancer;
KW skin disorder; neuropathic pain; inflammatory disorder;
KW coagulation diathesis; thrombosis; laundry detergent; skin care;
KW gene therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 31..52
FT /note= "hydrophobic transmembrane domain"
FT Active-site 202..203
FT /note= "residues spanning the zymogen cleavage site"
FT Active-site 243
FT /note= "catalytic triad residue"
FT Active-site 339
FT /note= "catalytic triad residue"
FT Misc-difference 361
FT /note= "Encoded by CTG"
FT Active-site 385
FT /note= "catalytic triad residue"
XX
PN WO200202011-A1.
XX
PD 10-JAN-2002.
XX
XX 08-JUN-2001; 2001WO-US18568.
XX
PR 30-JUN-2000; 2000US-0607745.
XX
PA (ORTH) ORTHO-MCNEIL PHARM INC.
XX
XX Darrow AL, Qi J, Andrade-Gordon P;
PI
XX
DR WPI; 2002-106601/14.
DR N-PSDB; AAI64284.
XX
PT Nucleic acid encoding a serine protease called D-G protein which is
PT useful for identifying modulators that are useful for treating a
PT condition which is mediated by protease D-G, e.g. cancer, skin
PT disorders, or neuropathic pain
XX
PS Claim 13; Fig 1B; 81pp; English.
PS
XX
CC The invention relates to an isolated and purified nucleic acid that
CC encodes a serine protease called D-G protein. The activity of the protein
CC of the invention may be described as cytostatic, antiinflammatory,
CC dermatological and anticoagulation. The serine protease of the invention
CC is a member of the trypsin/chymotrypsin-like (S1) serine protease family,
CC which play an important role in processes such as digestion and
CC regulatory amplification cascades through the proteolytic activation of
CC inactive zymogen precursors. Protease D-G modulating compounds are useful
CC for treating a condition which is mediated by protease D-G, e.g. cancer,
CC skin disorders, neuropathic pain, inflammatory disorders, or coagulation
CC diathesis/thrombosis. The polynucleotide encoding the protease is useful
CC for identifying, detecting or isolating mutant forms of DNA molecules
CC encoding the protease. The protease is useful for identifying modulators
CC of the functional protease. The D-G protein can be used for formulation
CC of compositions for laundry detergents and skin care products. Protease
CC D-G gene therapy may be used to introduce protease D-G into the cells of
CC target organisms. As the D-G protein is derived from a human, it is less
CC likely to produce an allergic reaction in sensitive individuals when used
CC in formulations for laundry detergents and skin care products. The
CC current sequence represents the human serine protease D-G amino acid
CC sequence.
XX
SQ Sequence 435 AA;

Query Match 100.0%; Score 2342; DB 23; Length 435;
Best Local Similarity 100.0%; Pred. No. 3.3e-194;

Matches	435;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0
1	MDPDSQPLNSLDVKPLRKPRIPMETFRKVGIPIIIALLSLASSIIIVVVLKIVLDKYFF	60							
1	MDPDSQPLNSLDVKPLRKPRIPMETFRKVGIPIIIALLSLASSIIIVVVLKIVLDKYFF	60							
61	LCCQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGN	120							
61	LCCQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGN	120							
121	WFSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGP	180							
121	WFSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGP	180							
181	CLSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLVT	240							
181	CLSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLVT	240							
241	AAHCFRKHTDVFNWKVRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLQFPPLTF	300							
241	AAHCFRKHTDVFNWKVRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLQFPPLTF	300							
301	SGTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAY	360							
301	SGTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAY	360							
361	QGEVTEKMMCAGIPEGVDTCQDSSGGLMYQSDQWHVVGIVSWGYGCGPSTPGVYTKV	420							
361	QGEVTEKMMCAGIPEGVDTCQDSSGGLMYQSDQWHVVGIVSWGYGCGPSTPGVYTKV	420							
421	SAYLNWIYNVWKAEL	435							
421	SAYLNWIYNVWKAEL	435							
ABU04912 standard; Protein; 435 AA.									
ABU04912;									
29-JAN-2003 (first entry)									
Human expressed protein tag (EPT) #1578.									
Translational profiling; expressed protein tag; EPT; kinase; phosphatase; protease; protease inhibitor; transporter; cytoskeletal protein; receptor; transcription factor; cancer; MHC; major histocompatibility complex; myeloma; colon cancer; gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.									
Homo sapiens.									
WO200278524-A2.									
10-OCT-2002.									
28-MAR-2002; 2002WO-US09671.									
28-MAR-2001; 2001US-279495P.									
21-MAY-2001; 2001US-292544P.									
08-AUG-2001; 2001US-310801P.									
01-OCT-2001; 2001US-326370P.									
04-DEC-2001; 2001US-336780P.									
20-FEB-2002; 2002US-358985P.									
(ZYCO-) ZYCOS INC.									
Chicz RM, Tomlinson AJ, Urban RG;									
WPI; 2003-040607/03.									

New polypeptides (e.g. kinases, phosphatases, proteases, transporters, cytoskeletal proteins, receptors or transcription factors), useful for treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or leukemia -

Example 2; SEQ ID No 1578; 134pp; English.

The invention describes a purified polypeptide, which comprises a fragment of a kinase, phosphatase, protease, protease inhibitor, transporter, cytoskeletal protein, receptor or transcription factor. The polypeptide is useful as an immunogenic composition for eliciting in a mammal an immunogenic response directed against any of the purified polypeptide. The purified polypeptide, or the antibody that binds to this polypeptide, is useful for treating cancer. The polypeptide is also useful for identifying compounds that binds to a naturally processed class I or class II MHC-binding polypeptide. The polypeptides and polynucleotides are particularly useful for treating or preventing myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma, lymphoma or leukaemia. These are also useful for screening agents for treating the above mentioned diseases. This sequence represents an expressed protein tag (EPT) isolated from human tissue for translational profiling.

Note: This sequence does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published/pct/sequences.

Query Match	100.0%; Score 2342; DB 24; Length 435;
Best Local Similarity	100.0%; Pred. No. 3.3e-194;
Matches 435; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MDPDSQPLNSLDVVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVWLKVIDKYFF 60
Db	
QY	1 MDPDSQPLNSLDVVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVWLKVIDKYFF 60
Db	
QY	61 LCGQLHFIPRKQLCDGELDCPLGDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGN 120
Db	
QY	61 LCGQLHFIPRKQLCDGELDCPLGDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGN 120
Db	
QY	121 WFSACFDNFTALAEACROMGYSSKPTFRAVEIGPDQDLVDVVEITENSQELMRNSSGP 180
Db	
QY	121 WFSACFDNFTALAEACROMGYSSKPTFRAVEIGPDQDLVDVVEITENSQELMRNSSGP 180
Db	
QY	181 CLSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLT 240
Db	
QY	181 CLSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLT 240
Db	
QY	241 AAHCFRKHTDVFNNWVRAGSDKLGSPFLAVAKIIIEFNPMYPKDNDIALMKLQFPPLTF 300
Db	
QY	241 AAHCFRKHTDVFNNWVRAGSDKLGSPFLAVAKIIIEFNPMYPKDNDIALMKLQFPPLTF 300
Db	
QY	301 SGTVRPICLPFFDEELTPATPLWIIIGWGFTKQNGKMSDILLQASVQVIDSTRCNADDAY 360
Db	
QY	361 QGEVTEKMMCAGIPEGGVDTCCQDSGGPLMYQSDQWHVVGVISWGYCGGPGSTPGVYTKV 420
Db	
QY	421 SAYLNWIYNVWKAEL 435
Db	
QY	421 SAYLNWIYNVWKAEL 435
Db	

RESULT 4

ABU04931

ID ABU04931 standard; Protein; 435 AA.

XX AC ABU04931;

XX AC

XX DT 29-JAN-2003 (first entry)

XX

DE Human expressed protein tag (EPT) #1597.

XX Translational profiling; expressed protein tag; EPT; kinase;

KW phosphatase; protease; protease inhibitor; transporter;

KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;

KW major histocompatibility complex; myeloma; colon cancer;

KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;

KW leukaemia.

OS Homo sapiens.

XX WO200278524-A2.

PN 10-OCT-2002.

PD 28-MAR-2002; 2002WO-US09671.

PF 28-MAR-2001; 2001US-279495P.

XX 21-MAY-2001; 2001US-292544P.

PR 08-AUG-2001; 2001US-310801P.

PR 01-OCT-2001; 2001US-326370P.

PR 04-DEC-2001; 2001US-336780P.

PR 20-FEB-2002; 2002US-358985P.

XX (ZYCO-) ZYCOS INC.

PA Chiciz RM, Tomlinson AJ, Urban RG;

XX WPI; 2003-040607/03.

DR New polypeptides (e.g. kinases, phosphatases, proteases, transporters,

XX cytoskeletal proteins, receptors or transcription factors), useful for

PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma

PT or leukemia .

XX Example 2; SEQ ID No 1597; 134pp; English.

PS The invention describes a purified polypeptide, which comprises a

CC fragment of a kinase, phosphatase, protease, protease inhibitor,

CC transporter, cytoskeletal protein, receptor or transcription factor.

CC The polypeptide is useful as an immunogenic composition for eliciting

CC in a mammal an immunogenic response directed against any of the purified

CC polypeptide. The purified polypeptide, or the antibody that binds to

CC this polypeptide, is useful for treating cancer. The polypeptide is

CC also useful for identifying compounds that binds to a naturally

CC processed class I or class II MHC-binding polypeptide. The polypeptides

CC and polynucleotides are particularly useful for treating or preventing

CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,

CC lymphoma or leukaemia. These are also useful for screening agents for

CC treating the above mentioned diseases. This sequence represents an

CC expressed protein tag (EPT) isolated from human tissue for translational

CC profiling.

CC Note: This sequence does not appear in the printed specification but was

CC obtained in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 435 AA;

SQ Query Match 100.0%; Score 2342; DB 24; Length 435;

Best Local Similarity 100.0%; Pred. No. 3.3e-194;

Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVVKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYIF 60

Db 1 MDPDSQPLNSLDVVKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYIF 60

QY 61 LCGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGN 120

Db 61 LCGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGN 120

QY 121 WFSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 180

Db 121 WFSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 180

QY 181 CLSGSLVSLHCLACGKSLKTPRVVGEEASVDSWPQVSIQYDKQHVCGSILDPHWVLT 240

Db 181 CLSGSLVSLHCLACGKSLKTPRVVGEEASVDSWPQVSIQYDKQHVCGSILDPHWVLT 240

QY 241 AAHCFRKHTDVFNWKVRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOFPLTF 300

Db 241 AAHCFRKHTDVFNWKVRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOFPLTF 300

QY 301 SGTVRPICLPFFDELTPTPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDA 360

Db 301 SGTVRPICLPFFDELTPTPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDA 360

QY 361 QGEVTEKMMCAGIPEGGVDTCQDGGPLMYQSDQHHVVGVISWGYCGGPGSTPGVYTKV 420

Db 361 QGEVTEKMMCAGIPEGGVDTCQDGGPLMYQSDQHHVVGVISWGYCGGPGSTPGVYTKV 420

QY 421 SAYLNWIYNVWKABL 435

Db 421 SAYLNWIYNVWKABL 435

RESULT 5

AAAY06437

ID AAY06437 standard; Protein; 435 AA.

XX AC AAY06437;

XX DT 27-SEP-1999 (first entry)

XX DE Human protease HUPM-6.

XX KW Serine protease; human; HUPM-6; cell proliferation; cancer;

XX OS immune disorder; inflammation; therapy.

XX OS Homo sapiens.

XX FH Key

FT Modified-site 26

FT Modified-site /note= "protein kinase C phosphorylation site"

FT Modified-site 59

FT Modified-site /note= "tyrosine kinase phosphorylation site"

FT Modified-site 93

FT Modified-site /note= "casein kinase II phosphorylation site"

FT Modified-site 128

FT Modified-site /note= "N-glycosylated"

FT Modified-site 144

FT Modified-site /note= "protein kinase C phosphorylation site"

FT Modified-site 148

FT Modified-site /note= "protein kinase C phosphorylation site"

FT Modified-site 176

FT Modified-site /note= "N-glycosylated"

FT Modified-site 197

FT Modified-site /note= "protein kinase C phosphorylation site"

FT Modified-site 200

FT Modified-site /note= "protein kinase C phosphorylation site"

FT Modified-site 231

FT Modified-site /note= "casein kinase II phosphorylation site"

FT Active-site 243

FT Modified-site /note= "characteristic of serine protease"

FT Modified-site 249

FT Modified-site /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"

FT Modified-site 260

FT Modified-site /note= "protein kinase C phosphorylation site"

FT Modified-site 303

FT Modified-site /note= "protein kinase C phosphorylation site"

FT Modified-site 351

FT Modified-site /note= "protein kinase C phosphorylation site"

FT Modified-site 360

FT Modified-site /note= "tyrosine kinase phosphorylation site"

FT Modified-site 365

FT Modified-site /note= "protein kinase C phosphorylation site"

FT Active-site 385 /note= "characteristic of serine protease"
XX WO9936550-A2.
PN 22-JUL-1999.
XX 12-JAN-1999; 99WO-US00655.
PF 16-JAN-1998; 98US-0008271.
PR (INCY-) INCYTE PHARM INC.
XX Bandman O, Corley NC, Guegler KJ, Hillman JL, Shah P;
PI Tang YT, Yue H;
XX WPI; 1999-430616/36.
DR N-PSDB; AAX87154.
XX Novel human protease molecules useful in the treatment of
PT developmental disorders and/or cancers
PT
XX Claim 1; Page 74-75; 90pp; English.
PS The present sequence represents novel human protease HUPM-6, as
XX deduced from the consensus sequence (see AAX87154) of overlapping
CC cDNA clones obtained from various libraries. Northern analysis
CC shows expression of HUPM-6 in gastrointestinal, and male and
CC female reproductive cDNA libraries. Approximately 65% of these
CC libraries are associated with neoplastic disorders, and 22% with
CC the immune response. The invention provides 12 new human
CC proteases, i.e. HUPM-1 to -12 (see AAX06432-43), and the
CC polynucleotides encoding them (see AAX87149-60). Also provided are
CC vectors, host cells and methods for producing HUPM polypeptides,
CC as well as agonists and antagonists of HUPM. Methods for treating
CC or preventing cell proliferative disorders and immune disorders
CC using HUPM or HUPM antagonists are claimed.
XX
SQ Sequence 435 AA;

Query Match 99.8%; Score 2338; DB 20; Length 435;
Best Local Similarity 99.8%; Pred. No. 7.3e-194;
Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLKVIDKYF 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||
1 MDPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLKVIDKYF 60
QY 61 LCGQPLHFIPRQKCDGELDCPLGEDEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGN 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||
61 LCGQPLHFIPRQKCDGELDCPLGEDEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGN 120
QY 121 WFSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGP 180
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||
121 WFSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGP 180
QY 181 CLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLT 240
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||
181 CLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLT 240
QY 241 AAHCFRKHDTVENWVKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFLTF 300
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||
241 AAHCFRKHDTVENWVKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFLTF 300
QY 301 SGTVRPICLPFFDEELTPATPLWIIIGWGTQKQNGKMSDILLQASVQVIDSTRCNADDDAY 360
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||
301 SGTVRPICLPFFDEELTPATPLWIIIGWGTQKQNGKMSDILLQASVQVIDSTRCNADDDAY 360
QY 361 QGEVTEKMMKAGIPGGGVDTCCQDSGGPLMYQSDQHVHVGVISWGYCGGSPTPGVYTKV 420
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||
361 QGEVTEKMMKAGIPGGGVDTCCQDSGGPLMYQSDQHVHVGVISWGYCGGSPTPGVYTKV 420
QY 421 SAYLNWIYNVWKAEL 435

Db |||||||||||||||||| 421 SAYLNWIYNVWKAEL 435
RESULT 6
ABU04930
ID ABU04930 standard; Protein; 435 AA.
XX AC ABU04930;
XX DT 29-JAN-2003 (first entry)
XX DE Human expressed protein tag (EPT) #1596.
XX KW Translational profiling; expressed protein tag; EPT; kinase;
KW phosphatase; protease; protease inhibitor; transporter;
KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer;
KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
KW leukaemia.
XX OS Homo sapiens.
XX PN WO200278524-A2.
XX PD 10-OCT-2002.
XX PF 28-MAR-2002; 2002WO-US09671.
XX PR 28-MAR-2001; 2001US-279495P.
PR 21-MAY-2001; 2001US-292544P.
PR 08-AUG-2001; 2001US-310801P.
PR 01-OCT-2001; 2001US-326370P.
PR 04-DEC-2001; 2001US-336780P.
PR 20-FEB-2002; 2002US-358985P.
XX (ZYCO-) ZYCOS INC.
XX Chicx RM, Tomlinson AJ, Urban RG;
WPI; 2003-040607/03.
PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
PT or leukemia
XX Example 2; SEQ ID No 1596; 134pp; English.
PS The invention describes a purified polypeptide, which comprises a
XX fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor.
CC The polypeptide is useful as an immunogenic composition for eliciting
CC in a mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to
CC this polypeptide, is useful for treating cancer. The polypeptide is
CC also useful for identifying compounds that binds to a naturally
CC processed class I or class II MHC-binding polypeptide. The polypeptides
CC and polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling.
CC Note: This sequence does not appear in the printed specification but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 435 AA;
SQ

Query Match 99.8%; Score 2338; DB 24; Length 435;
Best Local Similarity 99.8%; Pred. No. 7.3e-194;
Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVKPLRKPRIPMETFRKVGIPIIIIALLSLASIIIVVVLKIVLDKYYF 60
Db |||||
QY 61 LCGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGN 120
Db |||||
QY 121 WFSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 180
Db |||||
QY 181 CLSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCVGGSIIDPHWVLT 240
Db |||||
QY 241 AAHCFRKHTDVFNWVKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPPLTF 300
Db |||||
QY 301 SGTVRPICLPFFDEELTPATPLWIIGWGFTKQNGKMSDILLQASVQVIDSTRCNADDA 360
Db |||||
QY 361 QGEVTEKMCAGIPEGGVDFCQDSSGGPLMYQSDQWHVVGVISWGYCGGSPTPGVYTKV 420
Db |||||
QY 421 SAYLNWIYNVWKAE 435
Db |||||

RESULT 7
AAG79359
ID AAG79359 standard; Protein; 437 AA.
XX
AC AAG79359;
DT 21-AUG-2002 (first entry)
XX
DE CJA8 preferred sequence.
XX
KW Colorectal cancer; CGA7; CJA8; modulating protein; screening;
XX drug candidate; vaccine.
OS Homo sapiens.
XX
Key Location/Qualifiers
Domain 36..63
FT /note= "Transmembrane domain"
XX
PN US2002042067-A1.
XX
PD 11-APR-2002.
XX
PF 08-MAY-2001; 2001US-0851588.
XX
PR 17-AUG-2000; 2000US-0642252.
PR 06-SEP-2000; 2000US-0656002.
XX
PA (MACK/) MACK D.
PA (GISH/) GISH K C.
PA (WILS/) WILSON K E.
XX
PI Mack D, Gish KC, Wilson KE;
XX
DR WPI; 2002-453647/48.
DR N-PSDB; AAI72976.
XX
PT Screening drug candidates for treating colorectal cancer, comprises
PT determining the effect of the candidate on the expression profile gene

PT of CGA7 or CJA8 -
XX Disclosure; Fig 8; 40pp; English.
XX
CC The sequences given in AAG79356-59 show the colorectal cancer
CC proteins, CGA7 and CJA8. The CGA7 and CJA8 proteins are colorectal
CC cancer modulating proteins and have been mapped to chromosomes 2 (CGA7)
CC and 11 (CJA8). These sequences may be used in the method of the
CC invention for screening drug candidates. The method comprises adding
CC a drug candidate to a cell that expresses an expression profile gene
CC encoding CGA7, CJA8 or fragments and determining the effect of the
CC drug candidate on the expression of the expression profile gene. The new
CC methods are used to screen bioactive agents for the ability to bind to
CC or modulate the activity of CGA7 or CJA8 and evaluate the effect of a
CC candidate colorectal cancer drug. An antibody to CGA7 or CJA8 can inhibit
CC the activity of CGA7 or CJA8, respectively, and is used to screen for
CC an agent that can interfere with the binding of CGA7 or CJA8 to the
CC antibody. The antibody can be used to treat colorectal cancer. The
CC antibody or a fragment of it is used to localize a therapeutic group to
CC a colorectal cancer tissue, where the therapeutic group is a cytotoxic
CC agent or a radioisotope. Antisense molecules are used to inhibit
CC colorectal cancer in a cell. Nucleic acid segments encoding CGA7 or
CC CJA8 are used in a biochip. CGA7, CJA8 or a nucleic acid encoding it
CC are used to elicit an immune response. CGA7 or CJA8 is used to
CC determine the prognosis of an individual with colorectal cancer.
CC Nucleic acid encoding CGA7 or CJA8 can be used in vaccines.
XX
SQ Sequence 437 AA;

Query Match 99.8%; Score 2337; DB 23; Length 437;
Best Local Similarity 100.0%; Pred. No. 9e-194;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPIIIIALLSLASIIIVVVLKIVLDKYYFL 61
Db |||||
QY 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPIIIIALLSLASIIIVVVLKIVLDKYYFL 63
Db |||||
QY 62 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 121
Db |||||
QY 64 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 123
Db |||||
QY 122 FSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGPC 181
Db |||||
QY 124 FSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGPC 183
Db |||||
QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCVGGSIIDPHWVLT 241
Db |||||
QY 184 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCVGGSIIDPHWVLT 243
Db |||||
QY 242 AHCFRKHTDVFNWVKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPPLTF 301
Db |||||
QY 244 AHCFRKHTDVFNWVKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPPLTF 303
Db |||||
QY 302 GTVRPICLPFFDEELTPATPLWIIGWGFTKQNGKMSDILLQASVQVIDSTRCNADDA 361
Db |||||
QY 304 GTVRPICLPFFDEELTPATPLWIIGWGFTKQNGKMSDILLQASVQVIDSTRCNADDA 363
Db |||||
QY 362 GEVTEKMCAGIPEGGVDFCQDSSGGPLMYQSDQWHVVGVISWGYCGGSPTPGVYTKVS 421
Db |||||
QY 364 GEVTEKMCAGIPEGGVDFCQDSSGGPLMYQSDQWHVVGVISWGYCGGSPTPGVYTKVS 423
Db |||||
QY 422 AYLNWIYNVWKAE 435
Db |||||
QY 424 AYLNWIYNVWKAE 437
Db |||||

RESULT 8
ABU04915
ID ABU04915 standard; Protein; 437 AA.
XX
AC ABU04915;
XX
DT 29-JAN-2003 (first entry)

CC polypeptide. The purified polypeptide, or the antibody that binds to
CC this polypeptide, is useful for treating cancer. The polypeptide is
CC also useful for identifying compounds that binds to a naturally
CC processed class I or class II MHC-binding polypeptide. The polypeptides
CC and polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling.
CC Note: This sequence does not appear in the printed specification but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 437 AA;

Query Match 99.8%; Score 2337; DB 24; Length 437;
Best Local Similarity 100.0%; Pred. No. 9e-194;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYYFL 61
Db |||||
QY 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYYFL 63
Db |||||
QY 62 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 121
Db |||||
QY 64 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 123
Db |||||
QY 122 FSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 181
Db |||||
QY 124 FSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 183
Db |||||
QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 241
Db |||||
QY 184 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 243
Db |||||
QY 242 AHCFRKHTDVFVNWKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301
Db |||||
QY 244 AHCFRKHTDVFVNWKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 303
Db |||||
QY 302 GTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNAD DAYQ 361
Db |||||
QY 304 GTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNAD DAYQ 363
Db |||||
QY 362 GEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQWVHVGVISWGYGCGPSTPGVYTKVS 421
Db |||||
QY 364 GEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQWVHVGVISWGYGCGPSTPGVYTKVS 423
Db |||||
QY 422 AYLNWIYNVWKAE 435
Db |||||
QY 424 AYLNWIYNVWKAE 437
Db |||||

RESULT 10
ABU04935
ID ABU04935 standard; Protein; 437 AA.
XX
AC ABU04935;
XX
DT 29-JAN-2003 (first entry)
XX
DE Human expressed protein tag (EPT) #1601.
XX
KW Translational profiling; expressed protein tag; EPT; kinase;
KW phosphatase; protease; protease inhibitor; transporter;
KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer;
KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
leukaemia.
XX
OS Homo sapiens.
XX
PN WO200278524-A2.

XX 10-OCT-2002.
XX
XX 28-MAR-2002; 2002WO-US09671.
XX
XX 28-MAR-2001; 2001US-279495P.
PR 21-MAY-2001; 2001US-292544P.
PR 08-AUG-2001; 2001US-310801P.
PR 01-OCT-2001; 2001US-326370P.
PR 04-DEC-2001; 2001US-336780P.
PR 20-FEB-2002; 2002US-358985P.
XX
PA (ZYCO-) ZYCOS INC.
XX
XX Chicz RM, Tomlinson AJ, Urban RG;
XX WPI; 2003-040607/03.
DR
XX
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
PT or leukemia
XX
PS Example 2; SEQ ID No 1601; 134pp; English.
XX
CC The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor.
CC The polypeptide is useful as an immunogenic composition for eliciting
CC in a mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to
CC this polypeptide, is useful for treating cancer. The polypeptide is
CC also useful for identifying compounds that binds to a naturally
CC processed class I or class II MHC-binding polypeptide. The polypeptides
CC and polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling.
CC Note: This sequence does not appear in the printed specification but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 437 AA;

Query Match 99.8%; Score 2337; DB 24; Length 437;
Best Local Similarity 100.0%; Pred. No. 9e-194;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYYFL 61
Db |||||
QY 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYYFL 63
Db |||||
QY 62 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 121
Db |||||
QY 64 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 123
Db |||||
QY 122 FSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 181
Db |||||
QY 124 FSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 183
Db |||||
QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 241
Db |||||
QY 184 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 243
Db |||||
QY 242 AHCFRKHTDVFVNWKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301
Db |||||
QY 244 AHCFRKHTDVFVNWKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 303
Db |||||
QY 302 GTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNAD DAYQ 361
Db |||||
QY 304 GTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNAD DAYQ 363
Db |||||

QY 362 GEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQWHVVGVISWGYCGGPGSTPGVYTKVS 421
Db 364 GEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQWHVVGVISWGYCGGPGSTPGVYTKVS 423
QY 422 AYLNWIYNVWKAE 435
Db 424 AYLNWIYNVWKAE 437

RESULT 11

ABU04936
ID ABU04936 standard; Protein; 437 AA.
XX
AC ABU04936;
XX
DT 29-JAN-2003 (first entry)
DE Human expressed protein tag (EPT) #1602.
XX
KW Translational profiling; expressed protein tag; EPT; kinase;
KW phosphatase; protease; protease inhibitor; transporter;
KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer;
KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
KW leukaemia.

OS Homo sapiens.

XX

XX

PN WO200278524-A2.

XX

PD 10-OCT-2002.

XX

PF 28-MAR-2002; 2002WO-US09671.

XX

PR 28-MAR-2001; 2001US-279495P.

PR 21-MAY-2001; 2001US-292544P.

PR 08-AUG-2001; 2001US-310801P.

PR 01-OCT-2001; 2001US-326370P.

PR 04-DEC-2001; 2001US-336780P.

PR 20-FEB-2002; 2002US-358985P.

XX

PA (ZYCO-) ZYCOS INC.

XX

PI Chiciz RM, Tomlinson AJ, Urban RG;

XX

DR WPI; 2003-040607/03.

XX

PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,

XX

PS cytoskeletal proteins, receptors or transcription factors), useful for

XX

XX treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma

XX

XX or leukemia -

XX

XX Example 2; SEQ ID No 1602; 134pp; English.

CC

CC The invention describes a purified polypeptide, which comprises a

CC

CC fragment of a kinase, phosphatase, protease, protease inhibitor,

CC

CC transporter, cytoskeletal protein, receptor or transcription factor.

CC

CC The polypeptide is useful as an immunogenic composition for eliciting

CC

CC in a mammal an immunogenic response directed against any of the purified

CC

CC polypeptide. The purified polypeptide, or the antibody that binds to

CC

CC this polypeptide, is useful for treating cancer. The polypeptide is

CC

CC also useful for identifying compounds that binds to a naturally

XX

SQ Sequence 437 AA;

Query Match 99.8%; Score 2337; DB 24; Length 437;

Best Local Similarity 100.0%; Pred. No. 9e-194;

Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DPDSQPLNSLDVKPLRKPRIIPMETFRKVGIPPIIALLSLASIIIVVVLKIVLDKYFL 61

Db 4 DPDSQPLNSLDVKPLRKPRIIPMETFRKVGIPPIIALLSLASIIIVVVLKIVLDKYFL 63

QY 62 CGQPLHFIPRKQCDGELDCPLGEDEEHCVKSPFEGPAAVAVRLSKDRSTLQVLD SATGNW 121

Db 64 CGQPLHFIPRKQCDGELDCPLGEDEEHCVKSPFEGPAAVAVRLSKDRSTLQVLD SATGNW 123

QY 122 FSACFDNFTEALAEACRQMGYSKPTTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 181

Db 124 FSACFDNFTEALAEACRQMGYSKPTTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 183

QY 182 LSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWVLT 241

Db 184 LSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWVLT 243

QY 242 AHCFRKHTDVFNNKVRAGSKLGSFPLAVAKIIIEFNPMYPKDNIDALMKLOPPLTFS 301

Db 244 AHCFRKHTDVFNNKVRAGSKLGSFPLAVAKIIIEFNPMYPKDNIDALMKLOPPLTFS 303

QY 302 GTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 361

Db 304 GTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 363

QY 362 GEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQWHVVGVISWGYCGGPGSTPGVYTKVS 421

Db 364 GEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQWHVVGVISWGYCGGPGSTPGVYTKVS 423

QY 422 AYLNWIYNVWKAE 435

Db 424 AYLNWIYNVWKAE 437

RESULT 12

AAAY72559

ID AAAY72559 standard; Protein; 492 AA.

XX

AC AAAY72559;

XX

DT 02-MAY-2001 (first entry)

XX

DE Human seripancrin variant #1 protein.

XX

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX

XX arthritis; chronic obstructive pulmonary disorder; COPD; cancer;

XX

XX osteoporosis; aberrant wound healing; angiogenesis; diabetes;

XX

XX inflammatory disorder; stroke; cardiovascular disease; gene therapy;

XX

XX vaccine; cytostatic; cerebroprotective; vulnerary; osteopathic.

XX

OS Homo sapiens.

XX

XX WO200104141-A2.

XX

PD 18-JAN-2001.

XX

PF 04-JUL-2000; 2000WO-EP06211.

XX

PR 12-JUL-1999; 99EP-0113428.

XX

XX (MERE) MERCK PATENT GMBH.

XX

XX Suendermann B, Hofmann U, Matzku S, Wilbert O;

XX

DR WPI; 2001-147177/15.

XX

DR N-PSDB; AAD02557.

181 CLSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTF 240
181 CLSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTF 240
241 AAHCFRKHTDVFVNWKVRAGSKLGSFPFSLAVAKIIIEFNPMYPKNDIALMKLOFPLTF 300
241 AAHCFRKHTDVFVNWKVRAGSKLGSFPFSLAVAKIIIEFNPMYPKNDIALMKLOFPLTF 300
301 SGTVRPICLPFFDEELTPATPLWIIGWGFTKQGGKMSDILLQASVQVIDSTRCNADDAY 360
301 SGTVRPICLPFFDEELTPATPLWIIGWGFTKQGGKMSDILLQASVQVIDSTRCNADDAY 360
361 QGEVTEKMMKAGIPEGGVDTCCGDSGGPLMYQSDQNHVVGIVSWGVCYGGPSTPGVYTKV 420
361 QGEVTEKMMKAGIPEGGVDTCCGDSGGPLMYQSDQNHVVGIVSWGVCYGGPSTPGVYTKV 420
421 SAYLNWIYNVWK 432
421 SAYLNWIYNVWK 432
SULT 14
U04932
ABU04932 standard; Protein; 492 AA.
ABU04932;
29-JAN-2003 (first entry)
Human expressed protein tag (EPT) #1598.
Translational profiling; expressed protein tag; EPT; kinase;
phosphatase; protease; protease inhibitor; transporter;
cytoskeletal protein; receptor; transcription factor; cancer; MHC;
major histocompatibility complex; myeloma; colon cancer;
gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
leukaemia.
Homo sapiens.
WO200278524-A2.
10-OCT-2002.
28-MAR-2002; 2002WO-US09671.
28-MAR-2001; 2001US-279495P.
21-MAY-2001; 2001US-292544P.
08-AUG-2001; 2001US-310801P.
01-OCT-2001; 2001US-326370P.
04-DEC-2001; 2001US-336780P.
20-FEB-2002; 2002US-358985P.
(ZYCO-) ZYCOS INC.
Chicz RM, Tomlinson AJ, Urban RG;
WPI; 2003-040607/03.

CC this polypeptide, is useful for treating cancer. The polypeptide is
CC also useful for identifying compounds that binds to a naturally
CC processed class I or class II MHC-binding polypeptide. The polypeptides
CC and polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling.
CC Note: This sequence does not appear in the printed specification but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published pct sequences.

Sequence 492 AA;

Query Match 99.4%; Score 2329; DB 24; Length 492;
Best Local Similarity 100.0%; Pred. No. 5.1e-193;
Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY	1	MDPDSQPLNSLDVKPLRPRI	PMETFRKVGIPITIIALLSLASIIIVVWLIKVIIDKYIF	60
Dd	1	MDPDSQPLNSLDVKPLRPRI	PMETFRKVGIPITIIALLSLASIIIVVWLIKVIIDKYIF	60
QY	61	LCGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFPEGPAVAVRLSKDRSTLQVLDSATGN	120	
Dd	61	LCGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFPEGPAVAVRLSKDRSTLQVLDSATGN	120	
QY	121	WFSA CFDNFTALAEATA CRQM GYS SKPT FRAVEIGDPDQLDVVEITENSQELRMRNSSGP	180	
Dd	121	WFSA CFDNFTALAEATA CRQM GYS SKPT FRAVEIGDPDQLDVVEITENSQELRMRNSSGP	180	
QY	181	CLSGSLVSLHCLACGKSLKTPRVVVGEEASVDSPWPQVSIQYDKQHVCVCGSILDPHWILT	240	
Dd	181	CLSGSLVSLHCLACGKSLKTPRVVVGEEASVDSPWPQVSIQYDKQHVCVCGSILDPHWILT	240	
QY	241	AAHCFRKHTDFVNKWVRAGSKDLGSFPPLAVAKIIIEFNPMYPKDNDIALMKLOPPLTF	300	
Dd	241	AAHCFRKHTDFVNKWVRAGSKDLGSFPPLAVAKIIIEFNPMYPKDNDIALMKLOPPLTF	300	
QY	301	SGTVRPICLPFFDEELT PATPLWII GWGFT KONGKMSDILLQASQVIDSTRCNADDAY	360	
Dd	301	SGTVRPICLPFFDEELT PATPLWII GWGFT KONGKMSDILLQASQVIDSTRCNADDAY	360	
QY	361	QGEVTEKMWCAGIPEGGVDT CQGDSGGPLMYQS DQWHVVGI VSWGYCGGPGSTPGVYTKV	420	
Dd	361	QGEVTEKMWCAGIPEGGVDT CQGDSGGPLMYQS DQWHVVGI VSWGYCGGPGSTPGVYTKV	420	
QY	421	SAYINWIYNVWK	432	
Dd	421	SAYINWIYNVWK	432	

RESULT 15

AAE06931
ID AAE06931 standard; Protein: 437 AA.

AA
AC AAE06931;

DT 16-OCT-2001 (first entry)

DE Human membrane-type serine protease (MTSP) 3.

Human; transmembrane serine protease; membrane-type serine protease; MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic; lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy.

OS Homo sapiens.

xx	Key	Location/Qualifiers
FH	Domain	205.437
FT		

FT	cleavage-site	/label=	204	205
FT				

FT	Cleavage site	201
FT	Misc-difference	310

/note= "Unpaired cysteine"

WO200157194-A2.

09-AUG-2001.

02-FEB-2001; 2001WO-US03471.

03-FEB-2000; 2000US-0179982.

18-FEB-2000; 2000US-0183542.

22-JUN-2000; 2000US-0213124.

26-JUL-2000; 2000US-0220970.

08-SEP-2000; 2000US-0657986.
22-SEP-2000; 2000US-0234840

010101C70 0000007 / 00007 TTTT
TTTT

(CORV-) CORVAS INT INC.

Madison EL, Ong EO, Yeh J;

WPI; 2001-488877/53.

N-PSDB; AAD13114.

Novel single chain polypeptide comprising protease domain of type-II membrane-type serine protease or its catalytically active portion useful for treating and preventing cancer and tumor -

Claim 12; Page 199-200; 256pp; English.

The invention relates to transmembrane serine proteases and their corresponding nucleotides and the protease domain of a type-II membrane-type serine protease (MTSP). MTSP is useful for identifying compounds that modulate or inhibits its proteolytic activity and for formulating a medicament for treating neoplastic disease. MTSP and its corresponding nucleotides are useful in preventing or treating tumours or cancers such as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP is useful as a diagnostic marker for tumour development, growth and/or progression and as immunogens to generate antibodies that specifically bind to it. MTSP DNA is useful in a yeast two-hybrid system and in gene therapy. The present sequence is human MTSP3 protein

Sequence 437 AA;

Query Match 99.0%; Score 2319; DB 22; Length 437;

Best Local Similarity 99.5%; Pred. No. 3.2e-192;

```

Matches 432; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

2 DPDS DQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVVLKIVLDKYYFL 61

4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPIIALLSLASIIIVVVLKIVLDKYYFL 63

62 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLDSATGNW 121

64 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 123

122 FSACFDNFTEALAEACRQMGYSKPTFRAVEIGPDQDLDVEITENSQELMRNRSSGPC 181

124 FSACFDNFTEALAEACRQMGYSKPTFRAVEIGPDQDLVVEITENSQELMRNSSGPC 183

182 LSGSLVSLHCLACGKSLKTPRVVGEEASVDSWPWQVSIQYDKQHVC GGSILDPHWVLT A 241

184 LSGSLVSLHCLACGKSLKTPRVVGEEASVDSWPWQVSIQYDIQHVCSSILDPHWVLT 243

242 AHCFRKHTDVFNWKVRAGSDKLGSPSLAVAKIIIEFNPMYPKDNDIALMKLQFPLTFS 301

244 AHCFRKHTDVFNWKVRAGSDKLGSPSLAVAKIIIEFNPMYPKDNDIALMKLQPLTFS 303

302 GTVRPICLPFFDEELTPATPLWIIGWGFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 361

304 GTVRLICLPFFDEELTPATPLWIIGWGFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 363

362 GEVTEKMMCAGIPEGGVDTCCQDSSGGFLMYQSDQWHVVGIVSWGCGGPPSTPGVYTKVS 421

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 07:47:10 ; Search time 22.1389 Seconds
(without alignments)
831.351 Million cell updates/sec

Title: US-09-607-745-2
Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVKPLRKP.....VYTKVSAYLNWIYVWKAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2338	99.8	435	3	US-09-008-271A-6
2	2196	93.8	423	4	US-09-656-002-2
3	676.5	28.9	492	3	US-09-342-749-2
4	676.5	28.9	492	4	US-09-691-840-2
5	655.5	28.0	454	3	US-09-518-046-2
6	588.5	25.1	417	4	US-09-820-002-4
7	584	24.9	455	3	US-09-261-416-2
8	580	24.8	376	4	US-09-820-002-2
9	574	24.5	416	2	US-09-000-846-2
10	571	24.4	798	1	US-08-200-900A-2
11	571	24.4	798	5	PCT-US94-00616-2
12	568	24.3	418	1	US-08-508-448C-25
13	568	24.3	418	4	US-09-370-838-82
14	568	24.3	418	4	US-09-370-838-83
15	566	24.2	418	4	US-09-370-838-62
16	558.5	23.8	283	3	US-08-807-151-1
17	558.5	23.8	283	4	US-09-478-957-1
18	531.5	22.7	232	1	US-08-508-448C-19
19	512.5	21.9	256	2	US-09-027-337-3
20	512.5	21.9	256	4	US-09-644-600-3
21	511	21.8	638	2	US-08-681-151-3
22	507.5	21.7	255	3	US-08-944-483-67
23	503.5	21.5	285	4	US-09-023-942A-26
24	486	20.8	248	3	US-08-944-483-63
25	483	20.6	314	4	US-09-023-942A-6
26	482.5	20.6	284	4	US-09-387-375-7
27	478	20.4	314	3	US-09-008-271A-3

28	477	20.4	312	4	US-09-023-942A-4	Sequence 4, Appli
29	475	20.3	407	4	US-09-734-675-4	Sequence 4, Appli
30	472.5	20.2	317	4	US-09-386-629-7	Sequence 7, Appli
31	470	20.1	902	4	US-09-644-600-10	Sequence 10, Appli
32	467.5	20.0	290	4	US-09-386-653A-7	Sequence 7, Appli
33	466.5	19.9	405	4	US-09-734-675-2	Sequence 2, Appli
34	465.5	19.9	250	3	US-08-944-483-68	Sequence 68, Appli
35	463	19.8	855	2	US-09-027-337-2	Sequence 2, Appli
36	463	19.6	855	4	US-09-644-600-2	Sequence 2, Appli
37	458.5	19.6	316	4	US-09-387-375-9	Sequence 9, Appli
38	458.5	19.6	790	4	US-08-991-761A-13	Sequence 13, Appli
39	455.5	19.4	812	1	US-08-248-629A-1	Sequence 1, Appli
40	455.5	19.4	812	1	US-08-451-932-1	Sequence 1, Appli
41	455.5	19.4	812	1	US-08-452-260-1	Sequence 1, Appli
42	455.5	19.4	812	1	US-08-326-785-1	Sequence 1, Appli
43	455.5	19.4	812	2	US-08-612-788-1	Sequence 1, Appli
44	455.5	19.4	812	2	US-08-605-598B-1	Sequence 1, Appli
45	455.5	19.4	812	2	US-08-429-743-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-008-271A-6
; Sequence 6, Application US/09008271A
; Patent No. 6203979
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLNNOT13
; CLONE: 1337018
; SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
US-09-008-271A-6

Query Match 99.8%; Score 2338; DB 3; Length 435;
Best Local Similarity 99.8%; Pred. No. 6.2e-231;
Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVKPLRKPRIMETERKVGIPIIIIALLSLASIIIVVVLKIVLDKYIF 60
DB 1 MDPDSQPLNSLDVKPLRKPRIMETERKVGIPIIIIALLSLASIIIVVVLKIVLDKYIF 60

QY 61 LCGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGN 120
DB 61 LCGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGN 120

QY 121 WFSACDNFTEALAEACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 180
DB 121 WFSACDNFTEALAEACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 180

QY 181 CLSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLT 240
DB 181 CLSGSLVSLHCLACGESLKTTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLT 240

QY 241 AAHCFRKHDTVFNWVKRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTF 300
DB 241 AAHCFRKHDTVFNWVKRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTF 300

QY 301 SGTVPICLPFFDEELTPATPLIIGWGFTKQNGKMSDILLOASVQVIDSTRCNADDA 360
DB 301 SGTVPICLPFFDEELTPATPLIIGWGFTKQNGKMSDILLOASVQVIDSTRCNADDA 360

QY 361 QGEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGCGGPGSTPGVYTKV 420
DB 361 QGEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGCGGPGSTPGVYTKV 420

QY 421 SAYLNWIYNVWKAEL 435
DB 421 SAYLNWIYNVWKAEL 435

RESULT 2
US-09-656-002-2
; Sequence 2, Application US/09656002
; Patent No. 6455668
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-69108/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/656,002
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-656-002-2

Query Match 93.8%; Score 2196; DB 4; Length 423;
Best Local Similarity 97.1%; Pred. No. 2.1e-216;
Matches 408; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 16 PLRKPRIPMETFRKVGIPIIIIALLSLASIIIVVVLKIVLDKYIFLCGQPLHFIPRKQLC 75
DB 4 PCANPVPWRPSESIGIPIIIIIALLSLASIIIVVVLKIVLDKYIFLCGQPLHFIPRKQLC 63

QY 76 DGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNWFSACFDNFTEALAE 135
DB 64 DGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNWFSACFDNFTEALAE 123

QY 136 TACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGPCLSGSLVSLHCLACG 195
DB 124 TACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGPCLSGSLVSLHCLACG 183

QY 196 KSLKTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLTAAHCFRKHDTVFNWK 255
DB 184 KSLKTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLTAAHCFRKHDTVFNWK 243

QY 256 VRAGSDKLGSPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEE 315
DB 244 VRAGSDKLGSPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEE 303

QY 316 LTPATPLIIGWGFTKQNGKMSDILLOASVQVIDSTRCNADDAVQGEVTEKMMCAGIPE 375
DB 304 LTPATPLIIGWGFTKQNGKMSDILLOASVQVIDSTRCNADDAVQGEVTEKMMCAGIPE 363

QY 376 GGVDTCQDSDGGPLMYQSDQHVHVGIVSWGCGGPGSTPGVYTKVSAAYLNWIYNVWKAEL 435
DB 364 GGVDTCQDSDGGPLMYQSDQHVHVGIVSWGCGGPGSTPGVYTKVSAAYLNWIYNVWKAEL 423

RESULT 3
US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2

Query Match 28.9%; Score 676.5; DB 3; Length 492;
Best Local Similarity 39.1%; Pred. No. 1.2e-60;
Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;

QY 75 CDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNWFSACFDNFTEALA 134
DB 133 CDGVSHCPGGEDEENRCVRLY--GP-----NFILQVYSSQKSWHPVCQDDWNNYVG 181

QY 135 ETACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSG-----PCL 182
DB 182 RAACRDMGY--KNNFYSSQ-----GIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACS 232

QY 183 SGSLVSLHCLACGKSL--KTTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVL 239
DB 233 SKAVVSLRCIACGVNLNSSLRSQSRIVGGSALPGAWPQVSLHVQNVHVCSSILITPEWIV 292

QY 240 TAAHCFRKH-TDVFNWVKRAGSKLGSF----PSLAVAKIIIEFNPMY---PKNDIAL 291
DB 293 TAAHCFRKHPLNPNPWHWTAFAGILR-QSFMFYAGVQVEKVI---SHPNYDSKTKNDIAL 348

QY 292 MKLQPLTFSGTVRPICLPFFDEELTPATPLIIGWGFTKQNGKMSDILLOASVQVIDS 351
DB 349 MKLQPLTFNDLVKPVCLPNPGMWLQPEQLCWSISGWGATEEK-GKTSEVLNAKVLIET 407

QY 352 TRCNADDAVQGEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQ-WHVVGIVSWGCGG 410

Db 408 QRCNSRYVDNLITPAMICAGFLQGNVDSCQDGGPLVTSKNNIWLIGDTSWGSCKAK 467
QY 411 PSTPGVYTKVSAYLNWIYNVWKA 434
Db 468 AYRPGVYGNVMVFTDWIYRQMRAD 491

RESULT 4
US-09-691-840-2
; Sequence 2, Application US/09691840
; Patent No. 6444419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teph, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/691,840
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-691-840-2

Query Match 28.9%; Score 676.5; DB 4; Length 492;
Best Local Similarity 39.1%; Pred. No. 1.2e-60;
Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;
QY 75 CDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTAL 134
Db 133 CDGVSHCPGGEDENCVRLY--GP-----NFILQVYSSQRKSWHPVCQDDWNNY 181
QY 135 ETACRQMGYSKPTTFAVEIGPDQDLDVVEITENSQELMRNSSG-----PCL 182
Db 182 RAACRDMGY--KNNFYSSQ-----GIVDDSGSTSFMKLNTSAGNVDIYKKLYHSDACS 232
QY 183 SGLSVSLHCLACGKSL---KTPRVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVL 239
Db 233 SKAVVSLRCIACGVNLNLSRQSRIVGGEALPGAWPWQVSLHVQNVHVCGGSIIITPEWIV 292
QY 240 TAAHCFRKH-TDVFNWKVRAGSDKLGSF----PSLAVAKIIIEFNPMY---PKNDIAL 291
Db 293 TAAHCFVEKPLNNPWHWTAFAGILR-QSFMFYGAGYQVEKVI---SHENYDSKTKNDIAL 348
QY 292 MKLQPLTFSGTVRPICLPFFDEELTPATPLWIGWFTKQNGGKMSDILLQASVQVIDS 351
Db 349 MKLQKPLTFNDLVKPVCLPFGMWLQPEQLCWISGSGATEEK-GKTSEVLNAAKVLLIET 407
QY 352 TRCNADDAYQGEVTEKMCAGIPEGGVDTCCQDGGPLMYQSDQ-MHVVGIVSWGCGG 410
Db 408 QRCNSRYVDNLITPAMICAGFLQGNVDSCQDGGPLVTSKNNIWLIGDTSWGSCKAK 467
QY 411 PSTPGVYTKVSAYLNWIYNVWKA 434
Db 468 AYRPGVYGNVMVFTDWIYRQMRAD 491

RESULT 5
US-09-518-046-2
; Sequence 2, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.

; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: complete amino acid sequence of TADG-12
; OTHER INFORMATION: protein
US-09-518-046-2
Query Match 28.0%; Score 655.5; DB 3; Length 454;
Best Local Similarity 35.7%; Pred. No. 1.5e-58;
Matches 158; Conservative 73; Mismatches 149; Indels 63; Gaps 18;
QY 22 IPMETFRKVGIPILIALLSLASIIIVVVLIKVILD---KY----YFLCGQPLHFIPRKQL 74
Db 44 LPFEVFSQSSSLGIAL-----ILALAIGLGIHFDCSGKYRCRSSFKC---IELITR--- 92
QY 75 CDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTAL 134
Db 93 CDGVSDCKDGEDEYRC-----VRVGGQNAVLOVFTAA--SWKTMCSDDWKGHYA 139
QY 135 ETACRQMGYSS-----KPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPCL 182
Db 140 NVACAQLGFFPSYSSDNLRVSSLEGQFREFEVSIDHLLPDDKVTAHLHHSVYVREG---CA 196
QY 183 SGLSVSLHCLACG-KSLKTPRVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLTA 241
Db 197 SGHVVTLOCTACGHRRGYSRIVGGNMSLLSQWPWQASLOFQGYHLCGGSVITPLWIITA 256
QY 242 AHCERKHTDVF---NWKVAGSDKL--GSFPLAVAKIIIEFNPMY-PK--DNDIALMK 293
Db 257 AHCV---YDLYLPKSWTIQVGLVSLDNPAPSHLVEKIV---YHSKYKPKRLGNDIALMK 310
QY 294 LQFPLTFSGTVRPICLPFFDEELTPATPLWIGWFTKQNGGKMSDILLQASVQVIDSTR 353
Db 311 LAGPLTFNEMIQVCLPNSSENFPGDKVCWTSGWAT-EDGGDASPVLNHAAPLISNKI 369
QY 354 CNADDAYQGEVTEKMCAGIPEGGVDTCCQDGGPLMYQSDQ-MHVVGIVSWGCGGSPS 412
Db 370 CNHRDVGGIISPSMLCAGYLTGGVDSQDGGPLVCQERRLWKLVGATSPGIGCAEVN 429
QY 413 TPGVYTKVSAYLNWIYNVWKAEL 435
Db 430 KPGVYTRVTSFLDWIHEQMERDL 452

RESULT 6
US-09-820-002-4
; Sequence 4, Application US/09820002
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weiniu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001194
; CURRENT APPLICATION NUMBER: US/09/820,002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 417


```

; TYPE: PRT
; ORGANISM: HUMAN
US-09-820-002-4
Query Match      25.1%; Score 588.5; DB 4; Length 417;
Best Local Similarity 31.5%; Pred. No. 9.7e-52;
Matches 140; Conservative 60; Mismatches 161; Indels 83; Gaps 11;

QY 19 KPRIPMETFRKVGIPIIALLSLASIIIVVVLKIVLDKYFLCGQPLHFIPRKQLCDGE 78
Db 15 RPKVAALT---AGTLLLLTAIGAASWAIVAVLLR-----45

QY 79 LDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDNFTALAEATAC 138
Db 46 -----SDQE-----PLYPVQVSSADARLMVFDKTEGTWRLLCSSRSNARVAGLSC 90

QY 139 ROMGYSSKPTFRAVEIGPDQDLDVVEITENSQELMRNNSGGPCLSL 183
Db 91 EEMGF-----LRAL---THSELVTRTAGANGTSGFFCVDEGRPLPHTQRLLEVISPSDCPR 142

QY 184 GSLVSLHCLACG-KSLKTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLTAA 242
Db 143 GRFLAAICQDCGRKLPVDRIVGRDTSIGRWFPWQVSLRYDGAHLGGSLSGDWVLTAA 202

QY 243 HCFRKHDTVFN-WKVRAGSKDLGSPFSLAVAKIIIEFNPMY-----KNDIALMK 293
Db 203 HCFPENRVLNRVRFAGAVAGASPHGLQLGVQAVVYHGGYLPFRDPNSENSENIALVH 262

QY 294 LQFPLTFSGTVRPICLPFFDEELTPATPLWIIIGWFTKQGGKMSDILLQASVQVIDSTR 353
Db 263 LSSPLPLEYIOPVCLPAAGQALVDGKICTVTWGNT-QYGGQAGVLQEARVPIISNDV 321

QY 354 CNADDAYQGEVTEKMMKAGIPEGVDTCQDSDGGPLMYQ-----SDQWHVVGIVSWGVC 408
Db 322 CNGADFYGNQIKPKMFCAGYPEGGIDACQDSDGGPFVCEDSISRTPRWRLCGIVSWGTC 381

QY 409 GGPSTPGVYTKVSAYLNIWVNVWK 432
Db 382 ALAQKPGVYTKVSDFREWIFQAIK 405

RESULT 7
US-09-261-416-2
; Sequence 2, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TADG-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261,416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 2
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TADG-12 encoded by nucleotides
; OTHER INFORMATION: 144 to 1511 of Sequence 1
; Patent No. 6291663
US-09-261-416-2
Query Match      24.9%; Score 584; DB 3; Length 455;
Best Local Similarity 34.2%; Pred. No. 3.2e-51;
Matches 153; Conservative 73; Mismatches 150; Indels 72; Gaps 21;

QY 22 IPMETFRKVGIPIIALLSLASIIIVVVLKIVLD---KY-----YFLCGQPLHFIPRKQL 74
Db 44 LPFEVFSQSSSLGIIAL-----ILALAIGLGHFDSCGKYCRSSFKC---IELITR--- 92

QY 75 CDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDNFTALAE 134
```

```

Db 93 CDGVSDCKDGEDEYRC-----VRVGGQNAVLCQVFTAA--SWKTMCSDDWKGHYA 139
QY 135 ETACRQMGYSS-----KPTFRAVEIGPDQDLDVVEITENSQELMRNNSGGPCL 182
Db 140 NVACAQLGFPSPYVSSDNLRVSSLEGQFREEFVSIIDHLLPDDKVTALHHSVYVREG---CA 196
QY 183 SGLVSLHCLACG-KSLKTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLTAA 241
Db 197 SGHVVTLCQACGHRRGYSSRIVGNNMSLLSOWPWQASLOFQGYHLCGGSVITPLWIIITA 256
QY 242 AHCFRKHDTVF---NWKVRAGSDKL--GSFPSILAVAKIIIEFNPMY-PK--DNDIALMK 293
Db 257 AHCV---YDLPLPKSWTIQVGLVSLDNPAPSHLVEKIV---YHSKYKPKRLGNDIALMK 310
QY 294 LQFPLTFSGTVRPICLPFFDEELTPATPLWIIIGWFTKQGGKMSDILLQASVQVIDSTR 353
Db 311 LAGPLTFNEMIQPVCLPNSSENFDPDGKVCWTSWGAT-EDGGDASPVLNHAAPVLIS--- 366
QY 354 CNADDAYQGEVTEKMM-----CAGIPEG-GVDTCQDSDGGPLMYQSDQ-WHVVGIVSWGVC 407
Db 367 -NKDLQOPQGRVVRWHHLPLHALRGLPDGWRNNSCQDSDGGPLVCQERRLWKLVGATSFQIG 425
QY 408 CGPSTPGVYTKVSAYLNIWVNVKAE 435
Db 426 CADVKNKPGVYTRVTSFELDWIHEQMERDL 453

RESULT 8
US-09-820-002-2
; Sequence 2, Application US/098200002
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weinlu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001194
; CURRENT APPLICATION NUMBER: US/09/820,002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 376
; TYPE: PRT
; ORGANISM: HUMAN
US-09-820-002-2
Query Match      24.8%; Score 580; DB 4; Length 376;
Best Local Similarity 30.8%; Pred. No. 6.2e-51;
Matches 132; Conservative 57; Mismatches 146; Indels 94; Gaps 9;

QY 19 KPRIPMETFRKVGIPIIALLSLASIIIVVVLKIVLDKYFLCGQPLHFIPRKQLCDGE 78
Db 15 RPKVAALT---AGTLLLLTAIGAASWAIVAVLLR-----45

QY 79 LDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDNFTALAEATAC 138
Db 46 -----SDQE-----PLYPVQVSSADARLMVFDKTEGTWRLLCSSRSNARVAGLSC 90

QY 139 ROMGYSSKPTFRAVEIGPDQDLDVVEITENSQELMRNNSGGPCLSGSLVSLHCLACG-KS 197
Db 91 EEMGFSL-----DCPRGRFLAAICQDCGRK 116

QY 198 LKTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLTAAHCFRKHDTDFN-WKV 256
Db 117 LPVDRIVGRDTSIGRWFPWQVSLRYDGAHLGGSLSGDWVLTAAHCFPERNRVLSRWV 176

QY 257 RAGSDKLGSPFSLAVAKIIIEFNPMYP-----KNDIALMKLQFPLTFSGTVRPI 308
```

Db 177 FAGAVAQASPHGLQLGVQAVVYHGGYLPFRDPNSENNDIALVHLSPLPLTEYIQPVC 236
QY 309 LPFFDEELTPATPLWIGWFTKQNGGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKM 368
Db 237 LPAAGQALVDGKICTVTGWNT-QYYGQAGVQLQEARVPIISNDVCNGADFYGNQIKPKM 295
QY 369 MCAGIPEGGVDTCQDSSGGPLMYQ-----SDQHVHVGVISWGYGCGGFPSTPGVYTKVSAY 423
Db 296 FCAGYPEGGIDACQDSSGGPFVCEDSISRTPRNRLCGIVSWGTCALAQKPGVYTKVSDF 355
QY 424 LNWIYNVWK 432
Db 356 REWIFQAIK 364

RESULT 9
US-09-000-846-2
; Sequence 2, Application US/09000846
; Patent No. 5981830
; GENERAL INFORMATION:
; APPLICANT: WU, QINGYU
; APPLICANT: SADLER, JASPER
; TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH
; TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN, P.C.
; STREET: 2200 CLARENDON BLVD. SUITE 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: US
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,846
; FILING DATE: 30-DEC-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/866,058
; FILING DATE: 30-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: LEBOVITZ, RICHARD M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: BERLX 65P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-243-6333
; TELEFAX: 703-243-6410
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 416 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-000-846-2

Query Match 24.5%; Score 574; DB 2; Length 416;
Best Local Similarity 30.8%; Pred. NO. 3e-50;
Matches 135; Conservative 62; Mismatches 151; Indels 90; Gaps 10;

QY 30 VGPIIIALLSLASHIIWVLIKVLDKYVFLCGQLHFIPRKQLCDGELDCPLGEDEEH 89
Db 22 VGTLLFLTGTGIGAASWAIVTILLQ-----SDQE- 48
QY 90 CVKSFPEGPAVAVLSKDRSTLQVLDSATGNWFSACFDNFTEALAEATACROMGY----- 143
Db 49 -----PLYQVLSPGDSRLAVFDKTEGTWRLLCSSRSNARVAGLCCEMGFLRALAH 100
QY 144 -----SSKPTFRAVEIG-----PDQDLDVVEITENSQELMRNSSGCPCLSGSLVSL 189

Db 101 SELDVRTAGANGTSFFCVDEGGRLAQRLLDVISVCD-----CPRGRFLTA 147
QY 190 HCLACG-KSLKTPRVVGGEEASVDSWEPWQVSIQYDKQHVCGGSILDPHWVLTAAHCFRKH 248
Db 148 TCQDCGRRKLPVDRIVGGQDSSLRWEPWQVSLRYDGTGLCGGSLSGDWVLTAAHCFPER 207
QY 249 TDVFN-WKVRAGSDKLGSPFSLAVAKIIIEFNPMYP-----KNDIALMKLQFPLT 299
Db 208 NRVLRSRVRFAGAVARTSPHAVQLGVQAVIYHGGYLPFRDPTIDENSNDIALVHLSLSP 267
QY 300 FSGTVRPICLPFFDEELTPATPLWIIWGFTKQNGGKMSDILLQASVQVIDSTRCNADDA 359
Db 268 LTEYIQPVCLPAAQALVDGKVCVTVGWNT-QFYGQQAAMVLQEARVPIISNEVCNSPDF 326
QY 360 YQGEVTEKMMCAGIPEGGVDTCQDSSGGPLMYQ-----SDQHVHVGVISWGYGCGGPSTP 414
Db 327 YGNQIKPKMFCAGYPEGGIDACQDSSGGPFVCEDSISGTSRWRILCGIVSWGTCALARKP 386
QY 415 GVYTKVSAYLNWIYNVWK 432
Db 387 GVYTKVTDREWIFKAIK 404

RESULT 10
US-08-200-900A-2
; Sequence 2, Application US/08200900A
; Patent No. 5665566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,900A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinert, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-200-900A-2

Query Match 24.4%; Score 571; DB 1; Length 798;
Best Local Similarity 35.0%; Pred. No. 1.6e-49;
Matches 133; Conservative 67; Mismatches 138; Indels 42; Gaps 15;
QY 69 IPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDN 128
Db 435 IPLVNLCDGFPCHKDGSDEAHCVRLF-NGTTDSSGLVQPR--IQSI-----WHVACAEN 485
QY 129 FTEALAEATACROMGY---SSKPTFRAVEIGPDQDLDVVE-----ITENSQELMRNSSGP 180

Db 486 WTTQISDDVCQLLGLGTGNSSVPTF-STGGPPYVNLNTPANGSLILTPSQ----- 535
QY 181 CLSGSLVSLHC--LACGKSLKT-----PRVVGGEASVDSWPQVSIQYDKQHVCGGSILD 234
Db 536 CLEDSLILLCQNYKSCGKKLVTEQVSPKIVGSDSREGAWPWVVALYFDDQQVCGASLVS 595
QY 235 PHWVLTAAHC-FRKHTDVFNNKVRAGSKLGSF--PSLAVAKIIIEFNPMY---PKDND 288
Db 596 RDWLVSAAHCVYGRNMEPSKWKAUHLGLHMASNLTPQIETRLIDQIVINPHYNKRRKND 655
QY 289 IALMKLQFPPLTFSGTVRPICLPFPDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQV 348
Db 656 IAMHLEMKVNYTDYIQICLPEENQVFPFGPRICSIAGWGALIYQ-GSTADVLQEADVPL 714
QY 349 IDSTRCNADDAYQGEVTEKMMKAGIPEGGVDTCQDSGGPLMYQ-SDQWHVVGVISWGYG 407
Db 715 LSNEKCO-QQMPYENITENMVCAGYEAGVDSCQDGGPLMCQENNRWLLAGVTSFGYQ 773
QY 408 CGGPSTPGVYTKVSAYLNI 427
Db 774 CALPNRPGVIARVPRFTEWI 793

RESULT 11
PCT-US94-00616-2
; Sequence 2, Application PC/TUS9400616
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 33
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00616
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-00616-2

Query Match 24.4%; Score 571; DB 5; Length 798;
Best Local Similarity 35.0%; Pred. No. 1.6e-49;
Matches 133; Conservative 67; Mismatches 138; Indels 42; Gaps 15;

QY 69 IPRKQCDGELDCPLGEDEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDN 128
Db 435 IPLVNLCDGFPCHXKDGSDCAHCVRLF-NGTTDSSGLVQFR--IQSI-----WHVACAEN 485
QY 129 FTEALAEACRQMGY---SSKPTFRAVEIGPDQLDVVE---ITENSQELMRNSSGP 180
Db 486 WTTQISDDVCQLLGLGTGNSSVPTF-STGGPPYVNLNTPANGSLILTPSQ----- 535
QY 181 CLSGSLVSLHC--LACGKSLKT-----PRVVGGEASVDSWPQVSIQYDKQHVCGGSILD 234
Db 536 CLEDSLILLCQNYKSCGKKLVTEQVSPKIVGSDSREGAWPWVVALYFDDQQVCGASLVS 595
QY 235 PHWVLTAAHC-FRKHTDVFNNKVRAGSKLGSF--PSLAVAKIIIEFNPMY---PKDND 288
Db 596 RDWLVSAAHCVYGRNMEPSKWKAUHLGLHMASNLTPQIETRLIDQIVINPHYNKRRKND 655
QY 289 IALMKLQFPPLTFSGTVRPICLPFPDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQV 348
Db 656 IAMHLEMKVNYTDYIQICLPEENQVFPFGPRICSIAGWGALIYQ-GSTADVLQEADVPL 714
QY 349 IDSTRCNADDAYQGEVTEKMMKAGIPEGGVDTCQDSGGPLMYQ-SDQWHVVGVISWGYG 407

Db 715 LSNEKCO-QQMPYENITENMVCAGYEAGVDSCQDGGPLMCQENNRWLLAGVTSFGYQ 773
QY 408 CGGPSTPGVYTKVSAYLNI 427
Db 774 CALPNRPGVIARVPRFTEWI 793
RESULT 12
US-08-508-448C-25
; Sequence 25, Application US/08508448C
; Patent No. 5804410
; GENERAL INFORMATION:
; APPLICANT: Kazuyoshi YAMAKA et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
; TITLE OF INVENTION: TRYPSIN-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,448C
; FILING DATE: July 28, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-508-448C-25

Query Match 24.3%; Score 568; DB 1; Length 418;
Best Local Similarity 32.4%; Pred. No. 1.2e-49;
Matches 138; Conservative 79; Mismatches 149; Indels 60; Gaps 14;
QY 33 PIIIALLSLASHIIVVLKIVLDKYYFLCGQPLHFIPRK--QLCDGELDCPL----- 83
Db 16 PYVVCFIVAGVIVLAVTIALLV---YFLAFDQKSYFYRSSFQLLNVEYNSQLNSPATQE 72
QY 84 -----GEDEHCVKSPFEG-----PAVAVRLSKD-----RSTLQVLDSATGNWFSACFD 127
Db 73 YRTLSGRIESLITKTFKESNLNRNQFIRAHVAKLRQDGSVGRADVVMKFQFTRNNNGASM- 131
QY 128 NFEALAEACRQMGYSSKPTFRAVEIGPDQLDVVEITENSQELMRN--SSGPCLSGS 185
Db 132 ---KSRIESVLRQMLNS---GNLEINP---STEITSLTDQAAANWLINECGAGPDLI-- 180
QY 186 LVSLHCLACGKSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGGSILDPHWVLTAAHCF 245
Db 181 -----TLSEQRILGGTEAEGSWPQVSLRLNNAHHCGLSGLNNMWILTAHCF 229
QY 246 RKHTDVFNNKVRAGSKLGSFPL--AVAKIIIEFNPMYPKDNIALMKLQFLPFTFSGT 303

Db 230 RNSNPRDWIATSGIST--TFPKLRMRVRNLIHNHNYKSATHENDIALVRLNSVTFTKD 287
QY 304 VRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGE 363
Db 288 IHSVCLPAATQNIIPPGSTAYVTGWAQAEYAGHTVPE-LRQGVRIISNDVCNAPHSYNGA 346
QY 364 VTEKMCAGIPEGGVDTCCQDSGGGLMYQSDQ--WHVVGIVSWGYGCGGPSTPGVYTKVS 421
Db 347 ILSGMLCAGVPQGGVDACQDSGGPLVQEDSRRLWFIVGIVSWGDCQGLPKPGVYTRVT 406
QY 422 AYLNWI 427
Db 407 AYLDWI 412

RESULT 13
US-09-370-838-82
; Sequence 82, Application US/09370838
; Patent No. 644425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 82
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-82

Query Match 24.3%; Score 568; DB 4; Length 418;
Best Local Similarity 32.4%; Pred. No. 1.2e-49;
Matches 138; Conservative 79; Mismatches 149; Indels 60; Gaps 14;

QY 33 PIIIALLSLASHIIVVLIKVILDKYFLCGQPLHFIPRK--QLCDGELDCPL----- 83
Db 16 PYWCVFIVAGVILAVTIALLV---YFLAFDQKSYFYRSSFQLLNVEYNSQLNSPATQE 72
QY 84 -----GEDEHCVKSFPEG-----PAVAVRLSKD-----RSTLQVLDSATGNWFSACFD 127
Db 73 YRTLSGRIESLITKTFKESNLRNQFIRAHVAKLRQDGGVGRADVVMKFQFTRNNNGASM- 131
QY 128 NFTEALAEACRQMGYSSKPTTFAVEIGPDQDLDDVVEITENSQELMRN--SSGPCISSGS 185
Db 132 ---KSRIESVLRQMLNNS---GNLEINP---STEITSLTDQAAANWLINECGAGPDLI-- 180
QY 186 LVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLTAAHCF 245
Db 181 -----TLSEQRILGTEAEBSWPWQVSLRLNNAHCGGSLINNMWILTAAHCF 229
QY 246 RKHTDVFNMKVRAGSKLGSFPSL--AVAKIIIEFNPMYPKNDIALMKLQFPPLTFSGT 303
Db 230 RNSNPRDWIATSGIST--TFPKLRMRVRNLIHNHNYKSATHENDIALVRLNSVTFTKD 287
QY 304 VRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGE 363
Db 288 IHSVCLPAATQNIIPPGSTAYVTGWAQAEYAGHTVPE-LRQGVRIISNDVCNAPHSYNGA 346
QY 364 VTEKMCAGIPEGGVDTCCQDSGGGLMYQSDQ--WHVVGIVSWGYGCGGPSTPGVYTKVS 421
Db 347 ILSGMLCAGVPQGGVDACQDSGGPLVQEDSRRLWFIVGIVSWGDCQGLPKPGVYTRVT 406
QY 422 AYLNWI 427

Db 407 AYLDWI 412
RESULT 14
US-09-370-838-83
; Sequence 83, Application US/09370838
; Patent No. 644425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 83
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-83

Query Match 24.3%; Score 568; DB 4; Length 418;
Best Local Similarity 32.4%; Pred. No. 1.2e-49;
Matches 138; Conservative 79; Mismatches 149; Indels 60; Gaps 14;
QY 33 PIIIALLSLASHIIVVLIKVILDKYFLCGQPLHFIPRK--QLCDGELDCPL----- 83
Db 16 PYWCVFIVAGVILAVTIALLV---YFLAFDQKSYFYRSSFQLLNVEYNSQLNSPATQE 72
QY 84 -----GEDEHCVKSFPEG-----PAVAVRLSKD-----RSTLQVLDSATGNWFSACFD 127
Db 73 YRTLSGRIESLITKTFKESNLRNQFIRAHVAKLRQDGGVGRADVVMKFQFTRNNNGASM- 131
QY 128 NFTEALAEACRQMGYSSKPTTFAVEIGPDQDLDDVVEITENSQELMRN--SSGPCISSGS 185
Db 132 ---KSRIESVLRQMLNNS---GNLEINP---STEITSLTDQAAANWLINECGAGPDLI-- 180
QY 186 LVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLTAAHCF 245
Db 181 -----TLSEQRILGTEAEBSWPWQVSLRLNNAHCGGSLINNMWILTAAHCF 229
QY 246 RKHTDVFNMKVRAGSKLGSFPSL--AVAKIIIEFNPMYPKNDIALMKLQFPPLTFSGT 303
Db 230 RNSNPRDWIATSGIST--TFPKLRMRVRNLIHNHNYKSATHENDIALVRLNSVTFTKD 287
QY 304 VRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGE 363
Db 288 IHSVCLPAATQNIIPPGSTAYVTGWAQAEYAGHTVPE-LRQGVRIISNDVCNAPHSYNGA 346
QY 364 VTEKMCAGIPEGGVDTCCQDSGGGLMYQSDQ--WHVVGIVSWGYGCGGPSTPGVYTKVS 421
Db 347 ILSGMLCAGVPQGGVDACQDSGGPLVQEDSRRLWFIVGIVSWGDCQGLPKPGVYTRVT 406
QY 422 AYLNWI 427
Db 407 AYLDWI 412

RESULT 15
US-09-370-838-62
; Sequence 62, Application US/09370838
; Patent No. 644425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 07:49:15 ; Search time 27.5241 Seconds
(without alignments)
2885.234 Million cell updates/sec

Title: US-09-607-745-2
Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVKPLRKP.....VYTKVSAYLNWIYVWKAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 18259486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2338	99.8	435	12	US-10-180-719-6 Sequence 6, Appli
2	2337	99.8	437	9	US-09-851-588-8 Sequence 8, Appli
3	2319	99.0	437	11	US-09-776-191-4 Sequence 4, Appli
4	2297.5	98.1	432	11	US-09-888-257A-7 Sequence 7, Appli
5	2297.5	98.1	432	11	US-09-946-374-275 Sequence 275, App
6	2297.5	98.1	432	12	US-10-015-387A-275 Sequence 275, App
7	2297.5	98.1	432	12	US-10-063-735-112 Sequence 112, App
8	2297.5	98.1	432	12	US-10-006-130A-275 Sequence 275, App
9	2297.5	98.1	432	12	US-10-199-672-330 Sequence 330, App
10	2297.5	98.1	432	12	US-10-006-172A-275 Sequence 275, App
11	2297.5	98.1	432	12	US-10-187-749-330 Sequence 330, App
12	2297.5	98.1	432	12	US-10-194-457-330 Sequence 330, App
13	2297.5	98.1	432	12	US-10-184-642-330 Sequence 330, App
14	2297.5	98.1	432	12	US-10-196-747-330 Sequence 330, App
15	2297.5	98.1	432	12	US-10-015-392A-275 Sequence 275, App

16	2297.5	98.1	432	12	US-10-017-253A-275	Sequence 275, App
17	2297.5	98.1	432	12	US-10-173-689-330	Sequence 330, App
18	2297.5	98.1	432	12	US-10-173-690-330	Sequence 330, App
19	2297.5	98.1	432	12	US-10-173-691-330	Sequence 330, App
20	2297.5	98.1	432	12	US-10-173-692-330	Sequence 330, App
21	2297.5	98.1	432	12	US-10-173-694-330	Sequence 330, App
22	2297.5	98.1	432	12	US-10-173-698-330	Sequence 330, App
23	2297.5	98.1	432	12	US-10-173-699-330	Sequence 330, App
24	2297.5	98.1	432	12	US-10-173-707-330	Sequence 330, App
25	2297.5	98.1	432	12	US-10-174-569-330	Sequence 330, App
26	2297.5	98.1	432	12	US-10-174-583-330	Sequence 330, App
27	2297.5	98.1	432	12	US-10-174-587-330	Sequence 330, App
28	2297.5	98.1	432	12	US-10-174-589-330	Sequence 330, App
29	2297.5	98.1	432	12	US-10-174-591-330	Sequence 330, App
30	2297.5	98.1	432	12	US-10-175-736-330	Sequence 330, App
31	2297.5	98.1	432	12	US-10-175-742-330	Sequence 330, App
32	2297.5	98.1	432	12	US-10-175-744-330	Sequence 330, App
33	2297.5	98.1	432	12	US-10-175-745-330	Sequence 330, App
34	2297.5	98.1	432	12	US-10-175-748-330	Sequence 330, App
35	2297.5	98.1	432	12	US-10-175-751-330	Sequence 330, App
36	2297.5	98.1	432	12	US-10-175-754-330	Sequence 330, App
37	2297.5	98.1	432	12	US-10-176-480-330	Sequence 330, App
38	2297.5	98.1	432	12	US-10-176-489-330	Sequence 330, App
39	2297.5	98.1	432	12	US-10-176-754-330	Sequence 330, App
40	2297.5	98.1	432	12	US-10-176-755-330	Sequence 330, App
41	2297.5	98.1	432	12	US-10-176-759-330	Sequence 330, App
42	2297.5	98.1	432	12	US-10-176-920-330	Sequence 330, App
43	2297.5	98.1	432	12	US-10-176-922-330	Sequence 330, App
44	2297.5	98.1	432	12	US-10-176-924-330	Sequence 330, App
45	2297.5	98.1	432	12	US-10-176-984-330	Sequence 330, App

ALIGNMENTS

RESULT 1

US-10-180-719-6
; Sequence 6, Application US/10180719
; Publication No. US20030166246A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/180,719
; FILING DATE: 25-Jun-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271
; FILING DATE: 16-Jan-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555

Best Local Similarity 99.5%; Pred. No. 4.3e-224;
Matches 432; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYYFL 61
Db 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYYFL 63

QY 62 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFPEGPAVAVRLSKDRSTLQVLD SATGNW 121
Db 64 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFPEGPAVAVRLSKDRSTLQVLD SATGNW 123

QY 122 FSACFDNFTTEALAEATACROMGYSSKPTFRRAVEIGPDQDLVDVVEITENSQELMRNSSGPC 181
Db 124 FSACFDNFTTEALAEATACROMGYSSKPTFRRAVEIGPDQDLVDVVEITENSQELMRNSSGPC 183

QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 241
Db 184 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 243

QY 242 AHCFRKHTDVFVNWKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301
Db 244 AHCFRKHTDVFVNWKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 303

QY 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAVQ 361
Db 304 GTVRPICLPFFDEELTPATPLWIIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAVQ 363

QY 362 GEVTEKMCAGIPEGGVDTCCQDSGGPLMYQSDQWHVVGVISWGYGCGGPSTPGVYTKVS 421
Db 364 GEVTEKMCAGIPEGGVDTCCQDSGGPLMYQSDQWHVVGVISWGYGCGGPSTPGVYTKVS 423

QY 422 AYLNWIYNVWKAEL 435
Db 424 AYLNWIYNVWKAEL 437

RESULT 4

US-09-888-257A-7
; Sequence 7, Application US/09888257A
; Publication No. US20030060612A1
; GENERAL INFORMATION:
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Polakis, Paul
; APPLICANT: Smith, Victoria
; APPLICANT: Wood, William I.
; APPLICANT: Wu, Thomas D.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF TUMOR
; FILE REFERENCE: P5002R1
; CURRENT APPLICATION NUMBER: US/09/888,257A
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/063,540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: US 60/089,653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US 60/099,792
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/103,678
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: US 60/235,451
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: PCT/US99/20111
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: PCT/US00/04342
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/06666
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10
; SEQ ID NO 7
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-888-257A-7

Query Match 98.1%; Score 2297.5; DB 11; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYYFL 61
Db 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYYFL 63

QY 62 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFPEGPAVAVRLSKDRSTLQVLD SATGNW 121
Db 64 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFPEGPAVAVRLSKDRSTLQVLD SATGNW 123

QY 122 FSACFDNFTTEALAEATACROMGYSSKPTFRRAVEIGPDQDLVDVVEITENSQELMRNSSGPC 181
Db 124 FSACFDNFTTEALAEATACROMGYS-----RAVEIGPDQDLVDVVEITENSQELMRNSSGPC 178

QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 241
Db 179 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 238

QY 242 AHCFRKHTDVFVNWKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301
Db 239 AHCFRKHTDVFVNWKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 298

QY 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAVQ 361
Db 299 GTVRPICLPFFDEELTPATPLWIIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAVQ 358

QY 362 GEVTEKMCAGIPEGGVDTCCQDSGGPLMYQSDQWHVVGVISWGYGCGGPSTPGVYTKVS 421
Db 359 GEVTEKMCAGIPEGGVDTCCQDSGGPLMYQSDQWHVVGVISWGYGCGGPSTPGVYTKVS 418

QY 422 AYLNWIYNVWKAEL 435
Db 419 AYLNWIYNVWKAEL 432

RESULT 5

US-09-946-374-275
; Sequence 275, Application US/09946374
; Publication No. US20030073129A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2830P1C1
CURRENT APPLICATION NUMBER: US/09/946,374
CURRENT FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 60/098716
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098723
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098749
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098750
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098803
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098821
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098843
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/099536
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099602
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099642
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099741
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099754
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099763
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099792
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099808
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099815
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100388
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100584
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100661
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100662
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100664
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100683
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100684
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100710

PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100711
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100848
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100849
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100930
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101014
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101068
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101071
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101279
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: 60/101471
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101472
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101474
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101475
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101476
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101479
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101741
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101743
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101915
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101916
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/102207
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102240
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102307
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102330
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102331
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102484
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102487
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102570
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102571
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102684
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102687
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102965
PRIOR FILING DATE: 1998-10-02
PRIOR APPLICATION NUMBER: 60/103258
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103314
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103315
PRIOR FILING DATE: 1998-10-07

; PRIOR APPLICATION NUMBER: 60/103328
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103395
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103396
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103401
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103449
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103633
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103678
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103679
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103711
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/104257
; PRIOR FILING DATE: 1998-10-14
; PRIOR APPLICATION NUMBER: 60/104987
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105000
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105002
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105104
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 60/105169
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105266
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105693
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105694
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105807

Query Match 98.1%; Score 2297.5; DB 11; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 2 DPDSQPLNSLDVKPLRKPRIIPMETFRKVGIPIIALLSLASIIIVVVLKIVLDKYYFL 61
Db 4 DPDSQPLNSLDVKPLRKPRIIPMETFRKVGIPIIALLSLASIIIVVVLKIVLDKYYFL 63
QY 62 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAPAVRLSKDRSTLQVLD SATGNW 121
Db 64 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAPAVRLSKDRSTLQVLD SATGNW 123
QY 122 FSACFDNFTALAEATACRQMGYSKPTTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 181
Db 124 FSACFDNFTALAEATACRQMGYS-----RAVEIGPDQDLDVVEITENSQELMRNSSGPC 178
QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLT 241
Db 179 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLT 238
QY 242 AHCFRKHTDVFVNWKVRAGSKLGSFPFLAVAKIIIEFNPMYPKNDIALMKLOFPLTFS 301
Db 239 AHCFRKHTDVFVNWKVRAGSKLGSFPFLAVAKIIIEFNPMYPKNDIALMKLOFPLTFS 298
QY 302 GTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAVQ 361
Db 299 GTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAVQ 358
QY 362 GEVTEKMCAGIPEGGVDTCQDSDGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVS 421
Db 359 GEVTEKMCAGIPEGGVDTCQDSDGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVS 418
QY 422 AYLNWIYNVWKAE 435
Db 419 AYLNWIYNVWKAE 432

RESULT 7
US-10-063-735-112

RESULT 6
US-10-015-387A-275
; Sequence 275, Application US/10015387A
; Publication No. US20030135034A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2830P1C54
; CURRENT APPLICATION NUMBER: US/10/015,387A
; CURRENT FILING DATE: 2001-12-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 275
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-387A-275

Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 2 DPDSQPLNSLDVKPLRKPRIIPMETFRKVGIPIIALLSLASIIIVVVLKIVLDKYYFL 61
Db 4 DPDSQPLNSLDVKPLRKPRIIPMETFRKVGIPIIALLSLASIIIVVVLKIVLDKYYFL 63
QY 62 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAPAVRLSKDRSTLQVLD SATGNW 121
Db 64 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAPAVRLSKDRSTLQVLD SATGNW 123
QY 122 FSACFDNFTALAEATACRQMGYSKPTTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 181
Db 124 FSACFDNFTALAEATACRQMGYS-----RAVEIGPDQDLDVVEITENSQELMRNSSGPC 178
QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLT 241
Db 179 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLT 238
QY 242 AHCFRKHTDVFVNWKVRAGSKLGSFPFLAVAKIIIEFNPMYPKNDIALMKLOFPLTFS 301
Db 239 AHCFRKHTDVFVNWKVRAGSKLGSFPFLAVAKIIIEFNPMYPKNDIALMKLOFPLTFS 298
QY 302 GTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAVQ 361
Db 299 GTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAVQ 358
QY 362 GEVTEKMCAGIPEGGVDTCQDSDGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVS 421
Db 359 GEVTEKMCAGIPEGGVDTCQDSDGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVS 418
QY 422 AYLNWIYNVWKAE 435
Db 419 AYLNWIYNVWKAE 432

RESULT 7
US-10-063-735-112

; Sequence 112, Application US/10063735
; Publication No. US2003013882A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,735
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 112
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-735-112

Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

Qy 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVVLKIVLDKYYFL 61
Db 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVVLKIVLDKYYFL 63

Qy 62 CGQPLHFIPRKQKQCDGELDCPLGEDEEHCVKSFPEGPAAVAVRLSKORSTLQVLD SATGNW 121
Db 64 CGQPLHFIPRKQKQCDGELDCPLGEDEEHCVKSFPEGPAAVAVRLSKORSTLQVLD SATGNW 123

Qy 122 FSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPG 181
Db 124 FSACFDNFTALAEATACRQMGYS----RAVEIGPDQDLDVVEITENSQELMRNSSGPG 178

Qy 182 LSGSLVSLHCLACGKSLKTPRVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 241
Db 179 LSGSLVSLHCLACGKSLKTPRVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 238

Qy 242 AHCERKHTDVFENKVRAGSKLGSFPLSAVAKIIIEFNPMYPKNDIALMKLOFPLTFS 301
Db 239 AHCERKHTDVFENKVRAGSKLGSFPLSAVAKIIIEFNPMYPKNDIALMKLOFPLTFS 298

Qy 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQ 361
Db 299 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQ 358

Qy 362 GEVTEKMMCAGIPEGGVDTCQDSDGGPLMYQSDQWHVVGIVSWGVCYGGPSTPGVYTKVS 421
Db 359 GEVTEKMMCAGIPEGGVDTCQDSDGGPLMYQSDQWHVVGIVSWGVCYGGPSTPGVYTKVS 418

Qy 422 AYLNWIYNVWKAEL 435
Db 419 AYLNWIYNVWKAEL 432

RESULT 8
US-10-006-130A-275
; Sequence 275, Application US/10006130A
; Publication No. US20030148375A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman

; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2830PIC7
; CURRENT APPLICATION NUMBER: US/10/006,130A
; CURRENT FILING DATE: 2002-03-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 275
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-006-130A-275

Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

Qy 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVVLKIVLDKYYFL 61
Db 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVVLKIVLDKYYFL 63

Qy 62 CGQPLHFIPRKQKQCDGELDCPLGEDEEHCVKSFPEGPAAVAVRLSKORSTLQVLD SATGNW 121
Db 64 CGQPLHFIPRKQKQCDGELDCPLGEDEEHCVKSFPEGPAAVAVRLSKORSTLQVLD SATGNW 123

Qy 122 FSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPG 181
Db 124 FSACFDNFTALAEATACRQMGYS----RAVEIGPDQDLDVVEITENSQELMRNSSGPG 178

Qy 182 LSGSLVSLHCLACGKSLKTPRVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 241
Db 179 LSGSLVSLHCLACGKSLKTPRVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 238

Qy 242 AHCERKHTDVFENKVRAGSKLGSFPLSAVAKIIIEFNPMYPKNDIALMKLOFPLTFS 301
Db 239 AHCERKHTDVFENKVRAGSKLGSFPLSAVAKIIIEFNPMYPKNDIALMKLOFPLTFS 298

Qy 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQ 361
Db 299 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQ 358

Qy 362 GEVTEKMMCAGIPEGGVDTCQDSDGGPLMYQSDQWHVVGIVSWGVCYGGPSTPGVYTKVS 421
Db 359 GEVTEKMMCAGIPEGGVDTCQDSDGGPLMYQSDQWHVVGIVSWGVCYGGPSTPGVYTKVS 418

Qy 422 AYLNWIYNVWKAEL 435
Db 419 AYLNWIYNVWKAEL 432

RESULT 9
US-10-199-672-330
; Sequence 330, Application US/10199672
; Publication No. US20030148442A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/199,672
PRIOR FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: US/10/052,586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 330
LENGTH: 432
TYPE: PRT
ORGANISM: Homo Sapien
US-10-199-672-330

Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY	2	DPDSQPLNSLDVKPLRKPRIPIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYYFL	61
Db	4	DPDSQPLNSLDVKPLRKPRIPIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYYFL	63
QY	62	CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGNW	121
Db	64	CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGNW	123
QY	122	FSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC	181
Db	124	FSACFDNFTALAEATACRQMGYS-----RAVEIGPDQDLDVVEITENSQELMRNSSGPC	178
QY	182	LSGSLVSLHCLACGSLKTPRVVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTA	241
Db	179	LSGSLVSLHCLACGSLKTPRVVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTA	238
QY	242	AHCFRKHTDVFNWKVRAGSKLGSPPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS	301
Db	239	AHCFRKHTDVFNWKVRAGSKLGSPPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS	298
QY	302	GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVLDSTRCNADDAQY	361
Db	299	GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVLDSTRCNADDAQY	358
QY	362	GEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQVHVGVISWGYGCGPSTPGVYTKVS	421
Db	359	GEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQVHVGVISWGYGCGPSTPGVYTKVS	418
QY	422	AYLNWIYNVWKAEL	435
Db	419	AYLNWIYNVWKAEL	432

RESULT 10
US-10-006-172A-275

Sequence 275, Application US/10006172A
Publication No. US20030153000A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan I.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
ACIDS ENCODING THE SAME
FILE REFERENCE: P2830PIC11
CURRENT APPLICATION NUMBER: US/10/006,172A
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: 60/098716
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098723
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098749
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098750
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098803
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098821
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098843
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/099536
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099602
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099642
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099741
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099754
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099763
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099792
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099808
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099815
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100388
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100584
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100661

;
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100662
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100664
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100683
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100684
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100710
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100711
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100848
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100849
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100919
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100930
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/101014
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101068
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101071
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101279
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/101471
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101472
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101474
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101475
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101476
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101477
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101479
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101738
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101741
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101743
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101915
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101916
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/102207
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102240
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102307
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102330
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102331
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102484
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102487
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102570
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102571
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102684
; PRIOR FILING DATE: 1998-10-01

;
; PRIOR APPLICATION NUMBER: 60/102687
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/102965
; PRIOR FILING DATE: 1998-10-02
; PRIOR APPLICATION NUMBER: 60/103258
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103314
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103315
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103328
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103395
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103396
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103401
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103449
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103633
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103678
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103679
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103711
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/104257
; PRIOR FILING DATE: 1998-10-14
; PRIOR APPLICATION NUMBER: 60/104987
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105000
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105002
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105104
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 60/105169
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105266
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105693
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105694
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105807
; PRIOR FILING DATE: 1998-10-27
; PRIOR APPLICATION NUMBER: 60/105881
; PRIOR FILING DATE: 1998-10-27
; PRIOR APPLICATION NUMBER: 60/105882
; PRIOR FILING DATE: 1998-10-27
; PRIOR APPLICATION NUMBER: 60/106023
; PRIOR FILING DATE: 1998-10-28

Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-22;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

Qy 2 DPDS DQPLNSLDVKPLRKPRIPMETFRKVGIPIIALLSLASIIIVVVLKIVLDKYFL 61
Dd 4 DPDS DQPLNSLDVKPLRKPRIPMETFRKVGIPIIALLSLASIIIVVVLKIVLDKYFL 63
Qy 62 CGQPLHFIPRKQLCDGELDCPLGEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 121
Dd 64 CGQPLHFIPRKQLCDGELDCPLGEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 123
Qy 122 FSACFDNFTEALAEACRQMGYSKPTFRAVEIGPDQDLVDVVEITENSQELMRNSSGPC 181
Dd 124 FSACFDNFTEALAEACRQMGYS-----RAVEIGPDQDLVDVVEITENSQELMRNSSGPC 178
Qy 182 LSGSLVSLHCLACGKSLKTPRVVGEBASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 241

Db 179 LSGSLVSLHCLACGKSLKTRVVGGEASVDSWPWQVSIQYDKQHVCVCGSILDPHWVLT 238
Qy 242 AHCERKHTDVFNWVKRAGSDKLGSPPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
Db 239 AHCERKHTDVFNWVKRAGSDKLGSPPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 298
Qy 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 361
Db 299 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 358
Qy 362 GEVTEKMMCAGIPEGGVDTCQDSDGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVS 421
Db 359 GEVTEKMMCAGIPEGGVDTCQDSDGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVS 418
Qy 422 AYLNWIYNVWKAEL 435
Db 419 AYLNWIYNVWKAEL 432

RESULT 11
US-10-187-749-330
; Sequence 330, Application US/10187749
; Publication No. US20030153036A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/187,749
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: US/10/052,586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-187-749-330

Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.le-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;
Qy 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPIIALLSLASIIIVVLIKVILDKYYFL 61

Db 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPIIALLSLASIIIVVLIKVILDKYYFL 63
Qy 62 CGQPLHFIPRKQCDGELDCPLGEDEEHCVKSFPEGPAAVRLSKDESTLQVLD SATGNW 121
Db 64 CGQPLHFIPRKQCDGELDCPLGEDEEHCVKSFPEGPAAVRLSKDESTLQVLD SATGNW 123
Qy 122 FSACFDNTEALAEATACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 181
Db 124 FSACFDNTEALAEATACRQMGYS-----RAVEIGPDQDLDVVEITENSQELMRNSSGPC 178
Qy 182 LSGSLVSLHCLACGKSLKTRVVGGEASVDSWPWQVSIQYDKQHVCVCGSILDPHWVLT 241
Db 179 LSGSLVSLHCLACGKSLKTRVVGGEASVDSWPWQVSIQYDKQHVCVCGSILDPHWVLT 238
Qy 242 AHCERKHTDVFNWVKRAGSDKLGSPPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
Db 239 AHCERKHTDVFNWVKRAGSDKLGSPPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 298
Qy 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 361
Db 299 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 358
Qy 362 GEVTEKMMCAGIPEGGVDTCQDSDGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVS 421
Db 359 GEVTEKMMCAGIPEGGVDTCQDSDGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVS 418
Qy 422 AYLNWIYNVWKAEL 435
Db 419 AYLNWIYNVWKAEL 432

RESULT 12
US-10-194-457-330
; Sequence 330, Application US/10194457
; Publication No. US20030153037A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C296
; CURRENT APPLICATION NUMBER: US/10/194,457
; CURRENT FILING DATE: 2002-07-11
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28

; Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-194-457-330

```
Query Match      98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 2 DPDSQPLNSLDVKPLRKPRIIPMETFRKVGIPPIIIALLSLASIIIVVLIKVIIDKYFL 61
Db 4 DPDSQPLNSLDVKPLRKPRIIPMETFRKVGIPPIIIALLSLASIIIVVLIKVIIDKYFL 63

QY 62 CGQPLHFIIPRKQCDGELDCPLGDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 121
Db 64 CGQPLHFIIPRKQCDGELDCPLGDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 123

QY 122 FSACFDNFTEALAEACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 181
Db 124 FSACFDNFTEALAEACRQMGYS-----RAVEIGPDQDLDVVEITENSQELMRNSSGPC 178

QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHVLTA 241
Db 179 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHVLTA 238

QY 242 AHCFRKHTDVFNWKVRAGSKLGSFPPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301
Db 239 AHCFRKHTDVFNWKVRAGSKLGSFPPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 298

QY 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 361
Db 299 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 358

QY 362 GEVTEKMCAGIPEGGVDTCQDSSGGPLMYQSDQWHVVGIVSWGVCGGPSTPGVYTKVS 421
Db 359 GEVTEKMCAGIPEGGVDTCQDSSGGPLMYQSDQWHVVGIVSWGVCGGPSTPGVYTKVS 418

QY 422 AYLNWIYNVWKAE 435
Db 419 AYLNWIYNVWKAE 432
```

RESULT 13

```
US-10-184-642-330
; Sequence 330, Application US/10184642
; Publication No. US20030157635A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C194
; CURRENT APPLICATION NUMBER: US/10/184,642
; CURRENT FILING DATE: 2002-06-27
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-642-330
```

Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

```
QY 2 DPDSQPLNSLDVKPLRKPRIIPMETFRKVGIPPIIIALLSLASIIIVVLIKVIIDKYFL 61
Db 4 DPDSQPLNSLDVKPLRKPRIIPMETFRKVGIPPIIIALLSLASIIIVVLIKVIIDKYFL 63

QY 62 CGQPLHFIIPRKQCDGELDCPLGDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 121
Db 64 CGQPLHFIIPRKQCDGELDCPLGDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 123

QY 122 FSACFDNFTEALAEACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 181
Db 124 FSACFDNFTEALAEACRQMGYS-----RAVEIGPDQDLDVVEITENSQELMRNSSGPC 178

QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHVLTA 241
Db 179 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHVLTA 238

QY 242 AHCFRKHTDVFNWKVRAGSKLGSFPPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301
Db 239 AHCFRKHTDVFNWKVRAGSKLGSFPPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 298

QY 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 361
Db 299 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 358

QY 362 GEVTEKMCAGIPEGGVDTCQDSSGGPLMYQSDQWHVVGIVSWGVCGGPSTPGVYTKVS 421
Db 359 GEVTEKMCAGIPEGGVDTCQDSSGGPLMYQSDQWHVVGIVSWGVCGGPSTPGVYTKVS 418

QY 422 AYLNWIYNVWKAE 435
Db 419 AYLNWIYNVWKAE 432
```

RESULT 14

```
US-10-196-747-330
; Sequence 330, Application US/10196747
; Publication No. US20030162250A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C346
; CURRENT APPLICATION NUMBER: US/10/196,747
; CURRENT FILING DATE: 2002-07-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-196-747-330
```

Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

```
QY 2 DPDSQPLNSLDVKPLRKPRIIPMETFRKVGIPPIIIALLSLASIIIVVLIKVIIDKYFL 61
|||
```

Db 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPIIALLSLASIIIVVVLKIVLDKYYFL 63

QY 62 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGNW 121

Db 64 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGNW 123

QY 122 FSACFDNFTEALAEACFQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 181

Db 124 FSACFDNFTEALAEACFQMGYS-----RAVEIGPDQDLDVVEITENSQELMRNSSGPC 178

QY 182 LSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 241

Db 179 LSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 238

QY 242 AHCFRKHTDVENWVKRAGSKLGSFPPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301

Db 239 AHCFRKHTDVENWVKRAGSKLGSFPPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 298

QY 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAVQ 361

Db 299 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAVQ 358

QY 362 GEVTEKMMCAGIPEGGYDTCQDGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVS 421

Db 359 GEVTEKMMCAGIPEGGYDTCQDGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVS 418

QY 422 AYLNWIYNVWKAEL 435

Db 419 AYLNWIYNVWKAEL 432

RESULT 15

US-10-015-392A-275

; Sequence 275, Application US/10015392A

; Publication No. US20030166901A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Botstein, David

; APPLICANT: Desnoyers, Luc

; APPLICANT: Eaton, Dan I.

; APPLICANT: Ferrara, Napoleone

; APPLICANT: Fong, Sherman

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Grimaldi, Christopher J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Hillan, Kenneth J.

; APPLICANT: Pan, James

; APPLICANT: Paoni, Nicholas F.

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

; TITLE OF INVENTION: Acids Encoding the Same

; FILE REFERENCE: P2830PIC58

; CURRENT APPLICATION NUMBER: US/10/015,392A

; CURRENT FILING DATE: 2001-12-12

; PRIOR APPLICATION NUMBER: 60/098716

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098723

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098749

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098750

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098803

; PRIOR FILING DATE: 1998-09-02

; PRIOR APPLICATION NUMBER: 60/098821

; PRIOR FILING DATE: 1998-09-02

; PRIOR APPLICATION NUMBER: 60/098843

; PRIOR FILING DATE: 1998-09-02

; PRIOR APPLICATION NUMBER: 60/099536

; PRIOR FILING DATE: 1998-09-09

; PRIOR APPLICATION NUMBER: 60/099596

; PRIOR FILING DATE: 1998-09-09

; PRIOR APPLICATION NUMBER: 60/099598

; PRIOR FILING DATE: 1998-09-09

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 477

; SEQ ID NO 275

; LENGTH: 432

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-015-392A-275

Query Match 98.1%; Score 2297.5; DB 12; Length 432;

Best Local Similarity 98.8%; Pred. No. 6.1e-222;

Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPIIALLSLASIIIVVVLKIVLDKYYFL 61

Db 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPIIALLSLASIIIVVVLKIVLDKYYFL 63

QY 62 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGNW 121

Db 64 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGNW 123

QY 122 FSACFDNFTEALAEACFQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 181

Db 124 FSACFDNFTEALAEACFQMGYS-----RAVEIGPDQDLDVVEITENSQELMRNSSGPC 178

QY 182 LSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 241

Db 179 LSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTA 238

QY 242 AHCFRKHTDVENWVKRAGSKLGSFPPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301

Db 239 AHCFRKHTDVENWVKRAGSKLGSFPPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 298

QY 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAVQ 361

Db 299 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAVQ 358

QY 362 GEVTEKMMCAGIPEGGYDTCQDGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVS 421

Db 359 GEVTEKMMCAGIPEGGYDTCQDGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVS 418

QY 422 AYLNWIYNVWKAEL 435

Db 419 AYLNWIYNVWKAEL 432

Search completed: November 23, 2003, 07:54:47

Job time : 31.5241 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 23, 2003, 07:45:50 ; Search time 28.7208 Seconds
(without alignments)
1456.555 Million cell updates/sec

Title: US-09-607-745-2
Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVKPLRKP.....VYTKVSAYLNWIYVWKAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	584.5	25.0	417	1 S00845	hepsin (EC 3.4.21.
2	582.5	24.9	1019	1 A56318	enteropeptidase (E
3	571	24.4	1035	1 A43090	enteropeptidase (E
4	568.5	24.3	416	1 S33777	hepsin (EC 3.4.21.
5	552	23.6	1034	1 A53663	enteropeptidase (E
6	511	21.8	638	1 KQRTPL	plasma kallikrein
7	508	21.7	638	1 KQMSPL	plasma kallikrein
8	505	21.6	436	2 JX0172	acrosin (EC 3.4.21
9	504.5	21.5	1524	2 T30337	polyprotein - Afri
10	503	21.5	418	2 A37344	acrosin (EC 3.4.21
11	501.5	21.4	437	2 S18407	acrosin (EC 3.4.21
12	493.5	21.1	421	2 S29599	acrosin (EC 3.4.21
13	491.5	21.0	638	1 KQHUP	plasma kallikrein
14	488.5	20.9	1113	2 JE0315	low-density lipopr
15	481	20.5	421	1 S11674	acrosin (EC 3.4.21
16	478.5	20.4	415	1 A34170	acrosin (EC 3.4.21
17	474	20.2	855	2 JC7731	membrane-bound arg
18	470	20.1	761	2 JC5759	brain-specific ser
19	468.5	20.0	431	2 S47538	acrosin (EC 3.4.21
20	461.5	19.7	277	2 S35340	trypsin (EC 3.4.21
21	460.5	19.7	263	1 KYRTB	chymotrypsin (EC 3
22	458.5	19.6	267	2 S40006	trypsin (EC 3.4.21
23	458.5	19.6	790	1 PLPG	plasmin (EC 3.4.21
24	457.5	19.5	625	1 KFHU1	coagulation factor
25	455.5	19.4	812	1 PLMS	plasmin (EC 3.4.21
26	449	19.2	420	2 A55283	acrosin (EC 3.4.21
27	448	19.1	810	2 I46260	plasmin (EC 3.4.21
28	447.5	19.1	343	1 A57014	prostatin (EC 3.4.
29	445.5	19.0	455	2 A61545	plasmin (EC 3.4.21

ALIGNMENTS

RESULT 1

S00845
hepsin (EC 3.4.21.-) - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999
C;Accession: S00845
R;Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.
Biochemistry 27, 1067-1074, 1988
A;Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane doma
A;Reference number: S00845; MUID:88209431; PMID:2835076
A;Accession: S00845
A;Molecule type: mRNA
A;Residues: 1-417 <LEY>
A;Cross-references: EMBL:X07732; NID:G32063; PIDN:CAA30558.1; PID:G32064
C;Genetics:
A;Gene: GDB:HPN; TMRSS1; hepsin
A;Cross-references: GDB:I35685; OMIM:142440
A;Map position: 19q11-19q13.2
C;Superfamily: hepsin; trypsin homology
C;Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F;23-45/Domain: transmembrane #status predicted <TMN>
F;163-400/Domain: trypsin homology <TRY>
F;188-204,291-359,322-338,349-381/Disulfide bonds: #status predicted
F;203,257,353/Active site: His, Asp, Ser #status predicted

Query Match	25.0%;	Score 584.5;	DB 1;	Length 417;
Best Local Similarity	31.5%;	Pred. No. 7.2e-41;		
Matches 140;	Conservative 60;	Mismatches 161;	Indels 83;	Gaps 11;
QY	19	KPRIPMETFRKVGIPPIIIALLSLASIIIVVLLIKVILDKYIFLCQPLHFIPRKQLCDGE	78	
Db	15	RPKVAALT---AGTLLLLTAIGAASWAIVALLR-----	45	
QY	79	LDCPLGEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDNFTALAEATAC	138	
Db	46	-----SDQE-----PLYPVQVSSADARIMVFDKTEGTWELLCSSRSNARVAGLSC	90	
QY	139	ROMGYSSKPTFAVEIGPDQDLDVVEITEN-----SQELMRNSSSGPCLS	183	
Db	91	HEMGF-----LRAL---THSELDVTRTAGANGTSGFFCVDEGRLPHTQRLLEVISVCDQPR	142	
QY	184	GSLVSLHCLACG-KSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGGSILDPHWLTAA	242	
Db	143	GRFLAAICQDCGRKLPVDRIVGGRTSLGRWPQVSLRYDGAHLCCGSLLSGDWVLTAA	202	
QY	243	HCFRKHTDVFN-WKVRAGSDKLGSPFSLAVAKIIIEFNPMYP-----KONDIALMK	293	
Db	203	HCFPERNRVLSEWRVFAQAQASPHGLQLGVQAVVYHGGVLPFRDPNSENSENIALVH	262	
QY	294	LQFPLTFSGIVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTR	353	
Db	263	LSSPLPLTEYIQPVCLPAAGQALVDGKICTVTGWGNT-QYVGQAGVQLQEARVPIISNDV	321	

11801-1030/Domain: trypsin homology <TRY>
116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Binding
788-912,826-842,926-993,957-972,983-1011/Disulfide bonds: #status predicted
7841,892,987/Active site: His, Asp, Ser #status predicted

Query Match 24.4%; Score 571; DB 1; Length 1035;
Best Local Similarity 35.0%; Pred. No. 2.9e-39;
Matches 133; Conservative 67; Mismatches 138; Indels 42; Gaps 15;

69 IPRKQLCDGELDCPLGEDEEHCVKSFPEGPAPAVRLSKDRSTLQVLDSATGNWFSACFDN 128
672 IPLVNLCDGFPCHCKDGSDEAHCVRLP-NGTTDSSGLVQFR--IQSI-----WHVACAEN 722
129 FTEALAEACROMGY----SSKPTFRAVEIGPDQDLVVVE-----ITENSQELMRNSSGP 180
723 WTTQISDDVCQLLGLGTGNSVPTF-STGGGPYVNLNTAPNGSLILTPSQ----- 772
181 CLSGSLVSLHC--LACGKSLKT----PRVVGEEASVDSWPQVSIQYDKQHVCVGSILD 234
773 CLEDSLILLCQNYKSCGKLVTEVSPKIVGSDSREGAPWVVALYFDDQVCGASLVS 832
235 PHWVLTAAHC-FRKHTDVFNKVRAGSKLGSF--PSLAVAKIIIEFNPMY---PKDND 288
833 RDWLVSAAHCYVGRNMEPSKWKAVLGLHMASNLTSPOIETRLIDQVINPHYNKRKNND 892
289 IALMKLQFPLTFSGTVRPICLPFFDEELTPATPLWIIGWFTKQNGGKMSDILLQASVQV 348
893 IAMHLEMKVNTDYIQICLPENQVFPFPGRICSIAGWGALIYQ-GSTADVLQEADVPL 951
349 IDSTRCNADDAYQGEVTEKMCAGIPEGGVDTCCQDSSGGPLMYQ-SDQWHVVGVSWGYG 407
952 LSNEKCO-QOMPEYNITENMVCAGYEAGGVDSQCGDSGGPLMCQENRWLLAGVTSFGYQ 1010
408 CGGPSTPGVYTKVSAYLNWI 427
1011 CALPNRPGVIARVPRFTEWI 1030

RESULT 4
S33777
Species: Rattus norvegicus (Norway rat)
Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
Accession: S33777; S32013
R;Farley, D.; Reymond, F.; Nick, H.
Biochim. Biophys. Acta 1173, 350-352, 1993
Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.
Reference number: S33777; MUID:93305733; PMID:8318546
Accession: S33777
Status: preliminary
Molecule type: mRNA
Residues: 1-416 <FAR>
Cross-references: EMBL:X70900; NID:g57928; PIDN:CAA50256.1; PID:g57929
Superfamily: hepsin; trypsin homology
Keywords: hydrolase; liver; serine proteinase; transmembrane protein
22-44/Domain: transmembrane #status predicted <TMN>
162-399/Domain: trypsin homology <TRY>
187-203,290-358,321-337,348-380/Disulfide bonds: #status predicted
202,256,352/Active site: His, Asp, Ser #status predicted

Query Match 24.3%; Score 568.5; DB 1; Length 416;
Best Local Similarity 30.5%; Pred. No. 1.5e-39;
Matches 137; Conservative 64; Mismatches 155; Indels 93; Gaps 11;

19 KPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFLCGQLHFIPRKQLCDGE 78
14 RPKVAALT---VGTLLFLTIGIGASWAIVTILR----- 44
79 LDCPLGEDEEHCVKSFPEGPAPAVRLSKDRSTLQVLDSATGNWFSACFDNFTEALAEATAC 138
45 -----SDQE-----PLYQVQLSPGDSRLVLDKTEGTWRLLCSSRSNARVAGLGC 89
139 RQMGY-----SSKPTFRAVEIG-----PDQDLVDVEITENSQELMRNNS 178

Db 90 EEMGLRALAHSELVDVRTAGANGTSFFCVDEGGLPLAQRLLDVISVCD----- 138
QY 179 GPCLSGSLVSLHCLACG-KSLKTPRVVVGEEASVDSWPQVSIQYDKQHVCVGSILDPHW 237
Db 139 --CPGRFLTATCQDCGRKLPVDRIVGGQDSSLGRWPQVSLRYDGTGLCGSLLSGDW 196
QY 238 VLTAACHFRKHTDVFN-WKVRAGSDKLGSFPLSAVAKIIIEFNPMY-----KDND 288
Db 197 VLTAACHFPERNRVLNRVRFAGAVARTSPHAVQLGVQAVIYHGGYLPFRDPTIDENSND 256
QY 289 IALMKLQFPLTFSGTVRPICLPFFDEELTPATPLWIIGWFTKQNGGKMSDILLQASVQV 348
Db 257 IALVHLSLPLTEYIQVCLPAAQALVDGKVCVTVTGWGNT-QFYGQAVVQLQEARVPI 315
QY 349 IDSTRCNADDAYQGEVTEKMCAGIPEGGVDTCCQDSSGGPLMYQ-----SDQWHVVGVIS 403
Db 316 ISNEVCNSPDFYGNQIKPKMFCAGYPEGGIDACQDSSGGHFCVEDRISGTSRWRLCGIVS 375
QY 404 WGYGCGGPSTPGVYTKVSAYLNWIYNVWK 432
Db 376 WGTGICALARKPGVYTKVIDFREWIFQAIAK 404

RESULT 5
A53663
enteropeptidase (EC 3.4.21.9) precursor [validated] - pig
N;Alternate names: enterokinase
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
C;Accession: A53663
R;Matsumura, M.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokawa, F.
J. Biol. Chem. 269, 19976-19982, 1994
Title: Structural characterization of porcine enteropeptidase.
Reference number: A53663; MUID:94327548; PMID:8051081
Accession: A53663
Molecule type: mRNA
Residues: 1-1034 <MAT>
Cross-references: GB:D30799; NID:g505122; PIDN:BAA06459.1; PID:g505123
Note: parts of this sequence, including the amino ends of three chains isolated from the
Comment: The mechanism of association with the membrane of the intestinal brush border
ated below) or with amino-terminal myristoylation of the heavy chain.
Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involve
C;Function:
Description: cleaves activation peptide from trypsinogen to produce active trypsin
A;Pathway: intestinal digestive hydrolase cascade
C;Superfamily: enteropeptidase; Ctr/Cls repeat homology; LDL receptor ligand-binding rep
C;Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F;22-38/Domain: transmembrane #status predicted <TMN>
F;52-117/Product: enteropeptidase mini chain #status predicted <MCH>
F;118-799/Product: enteropeptidase heavy chain #status predicted <HCH>
F;199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;357-519/Domain: NAM homology <NAM>
F;541-646/Domain: Ctr/Cls repeat homology <CTR>
F;658-692/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRCF
F;800-1034/Product: enteropeptidase light chain #status predicted <LCH>
F;800-1029/Domain: trypsin homology <TRY>
F;116,147,170,194,283,343,350,403,455,485,518,549,645,697,701,721,740,761,804,863,902,964
F;787-911,825-841,925-992,956-971,982-1010/Disulfide bonds: #status predicted
F;840,891,986/Active site: His, Asp, Ser #status predicted

Query Match 23.6%; Score 552; DB 1; Length 1034;
Best Local Similarity 34.2%; Pred. No. 1.1e-37;
Matches 127; Conservative 69; Mismatches 141; Indels 34; Gaps 14;

QY 74 LDCGELDCPLGEDEEHCVKSFPEGPAPAVRLSKDRSTLQVLDSATGNWFSACFDNFTEAL 133
Db 676 LDCGFSHCKDGSDEAHCVR-FLNGTANNGLVQFR--IQSI-----WHTACAENWTQT 726
QY 134 AETACRQMGY---SSKPTFRAVEIGPDQDLVDVEITENSQELMRNNSGPGCLSGSLVSL 189

Db 727 SDDVCQLGLGTGNSMPPFSSGG-GP-----FVKLNTAPNGSLILTASEQCFFEDSLILL 780

QY 190 HC--LACGKSL-----KTPRVVVGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTAH 243

Db 781 QCNHSCGKKQVAQEVSPKIVGNDREGAWPWWVALYNGQLCCASLVRDLVSAH 840

QY 244 C-FRKHTDVFNKVRAGSKLGSF--PSLAVAKIIIEFNPMY---PKNDIALMKLQFP 297

Db 841 CVYGRNLEPSKWKAILGLHMTSNLTSPQIVTRLIDEIVINPHYNRRRKSDIAMHLEFK 900

QY 298 LTFSGTVRPICLPFFDEBLTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNAD 357

Db 901 VNYTDYIOPICLPEENQVPPGGRICSIAGWGVYQGSP-ADILQADVPLLSNEKCO-Q 958

QY 358 DAYQGEVTEKVMCAGIPEGGVDTCCQDGGPLM-YQSDQWVVGVISWGYCGGPSTPGV 416

Db 959 QMPEYNITENMMCMAGVEEGIDSCQDGGPLMCLNNRWLLAGVTSFGYQCALPNRPGV 1018

QY 417 YTKVSAYLNIWI 427

Db 1019 YARVPKFTTEWI 1029

RESULT 6

KQRTPL

plasma kallikrein (EC 3.4.21.34) precursor - rat

N;Alternate names: Fletcher factor; kininogenin; serum kallikrein

C;Species: Rattus norvegicus (Norway rat)

C;Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 18-Jun-1999

C;Accession: A39180; A33320; S06851; I53041; S06852

R;Beaubien, G.; Rosinski-Chupin, I.; Mattei, M.G.; Mbikay, M.; Chretien, M.; Seidah, N.G. Biochemistry 30, 1628-1635, 1991

A;Title: Gene structure and chromosomal localization of plasma kallikrein.

A;Reference number: A39180; MUID:91129236; PMID:1993180

A;Accession: A39180

A;Molecule type: DNA

A;Residues: 1-638 <BEA>

A;Cross-references: GB:J05315

A;Note: the authors translated the codon GAG for residue 81 as Gln

R;Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazur DNA 8, 563-574, 1989

A;Title: The cDNA structure of rat plasma kallikrein.

A;Reference number: A33320; MUID:90091743; PMID:2598771

A;Accession: A33320

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-638 <SEI>

A;Cross-references: GB:M30282; NID:G205010; PIDN:AAA41463.1; PID:G205011

A;Note: part of this sequence, including the amino ends of both the heavy and light chain

R;Paquin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G. Biochim. Biophys. Acta 999, 103-110, 1989

A;Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development

A;Reference number: S06851; MUID:90089457; PMID:2597701

A;Accession: S06851

A;Molecule type: protein

A;Residues: 20-45;391-413 <PAQ>

R;Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; Lazur DNA Cell Biol. 8, 563-574, 1989

A;Title: The cDNA structure of rat plasma kallikrein.

A;Reference number: I53041

A;Accession: I53041

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-638 <RES>

A;Cross-references: GB:M58590; NID:G206721; PIDN:AAA42069.1; PID:G206722

C;Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w

C;Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a li

are linked by one or more disulfide bonds.

C;Genetics:

A;Gene: PK

C;Superfamily: coagulation factor XI; trypsin homology

C;Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-390/Product: plasma kallikrein heavy chain #status experimental <MAT1>

F;20-109/Domain: apple repeat <AP1>

F;110-199/Domain: apple repeat <AP2>

F;200-289/Domain: apple repeat <AP3>

F;291-380/Domain: apple repeat <AP4>

F;391-638/Product: plasma kallikrein light chain #status experimental <MAT2>

F;391-621/Domain: trypsin homology <TRY>

F;21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-

F;127,215,308,453,459,494/Binding site: carbohydrate (Asn) (covalent) #status predicted

F;396/Binding site: carbohydrate (Asn) (covalent) #status experimental

F;434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 21.8%; Score 511; DB 1; Length 638;

Best Local Similarity 32.0%; Pred.No. 1.6e-34;

Matches 128; Conservative 62; Mismatches 106; Indels 104; Gaps 17;

QY 64 QPLHFIPRKQLCDG-----ELDCPL--GED--EEHCVKSF-----PEGP 98

Db 290 EPCHF----KIYSGVAFEGEELNATFVGADACQETCTKTRCQFFTYSLLPQDCKAEGC 345

QY 99 AVAVRLSKDRSTLQVLDSATGNWFSACFDNFTEALAEACQMGYSKPTFRAVEIGPDQ 158

Db 346 KCSLRLSTDGSPTRITYEAQGS-----SGYS----- 371

QY 159 DLDVVEITENSQELMRNNSGPGCLSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPQV 218

Db 372 -LRLCKVVESSD-----CTTKINA-RIVGGTNSSLGEPWQV 406

QY 219 SIQ---YDKQHVCGGSILDPHWLTAHCFR--KHTDVFNWVKVRAG---SDKLGSPFSL 269

Db 407 SLQKLVSNQNMCGGSIIGROWLTAHCFDGPYPDV--WRIYGGILNLSEITNKTFFS 464

QY 270 AVAKIIIEFNPMYPKNDIALMKLQFPPLTFSGTVRPICLPFFDEBLTPATPLWIGWF 329

Db 465 SIKELIIHOKYKMSSEGSYDIALIKLQTPLNYTEFQKPICLPSKADTNTIYTNCWVTGWGY 524

QY 330 TKQNGKMSDILLQASVQVIDSTRCNADDAYQGEV-TEKMMCAGIPEGGVDTCCQDSGCP 388

Db 525 TKER-GETONILQKATIPLVNPEC--QKKYRDYVITKQMICAGYKEGGIDACKGDSGCP 581

QY 389 LMYQ-SDQWHVVGVISWGYCGGPGSTPGVYTKVSAYLNIWI 427

Db 582 LVCKHSGRWQLVGITSWGEGCARKEQPGVYTKVAEYIDWI 621

RESULT 7

KOMSP

plasma kallikrein (EC 3.4.21.34) precursor - mouse

C;Species: Mus musculus (house mouse)

C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999

C;Accession: A36557

R;Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochemont DNA Cell Biol. 9, 737-748, 1990

A;Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and compariso

A;Reference number: A36557; MUID:91030844; PMID:2264928

A;Accession: A36557

A;Molecule type: mRNA

A;Residues: 1-638 <SEI>

A;Cross-references: GB:M58588; NID:G200358; PIDN:AAA63393.1; PID:G200359

A;Note: part of this sequence, including the amino ends of both the heavy and light chain

C;Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w

C;Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a li

are linked by one or more disulfide bonds.

C;Superfamily: coagulation factor XI; trypsin homology

C;Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-390/Product: plasma kallikrein heavy chain #status experimental <HCH>

F;20-109/Domain: apple repeat <AP1>

F;110-199/Domain: apple repeat <AP2>

F;200-289/Domain: apple repeat <AP3>

F;291-380/Domain: apple repeat <AP4>

F;391-638/Product: plasma kallikrein light chain #status experimental <LCH>

F;391-621/Domain: trypsin homology <TRY>

F;21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322
F;127,215,308,396,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 21.7%; Score 508; DB 1; Length 638;
Best Local Similarity 41.0%; Pred. No. 2.9e-34;
Matches 112; Conservative 45; Mismatches 82; Indels 34; Gaps 10;

QY 174 MRNSSGPCLSGSLVSLHCLACGKSLKTP-----RVVGEEASVDSWPWQVSIQ---Y 222
Db 364 MQGSSG-----YSLRLCKLVDSPTCTTKINARIVGNTASLGEWPWQVSLQVKLV 413
QY 223 DKQHVCGGSILDPHWLTAACHFR--KHTDVFENWVKVRAG---SDKLGSFPPSLAVAKIII 276
Db 414 SQTHLCGGSIIIGRQWVLTAAHCFDGIYPDV--WRIYGGILSLSEITKBTSPSSRIKELII 471
QY 277 IEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPATPLMIIGWFTKQNGGK 336
Db 472 HQEYKVSEGYDIALIKLQTPLNYTEFQKPICLPSKADTNTIYTNVCWVTGWYTKBQ-GE 530
QY 337 MSDILLQASVQVIDSPRCNADDAAYQGEVTEK-MMCAGIPEGGVDTCQDSDGGPLMYQ-SD 394
Db 531 TQNILQKATIPLVNPEC--QKKYRDYVINKQMICAGYKEGGTDACKDSDGGPLVCKHSG 588
QY 395 QWHVVGIVSWGCGGSPSTPGVYTKVSAYLNWI 427
Db 589 RWQLVGITSWGEGCGKDKQPGVYTKVSEYMDWI 621

RESULT 8
JX0172
acrosin (EC 3.4.21.10) precursor form 3 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Jun-2000
C;Accession: JX0172; JX0138
R;Watanabe, K.; Baba, T.; Kashiwabara, S.; Okamoto, A.; Arai, Y.
J. Biochem. 109, 828-833, 1991
A;Title: Structure and organization of the mouse acrosin gene.
A;Reference number: JX0172; MUID:92041732; PMID:1939002

A;Accession: JX0172
A;Molecule type: DNA
A;Residues: 1-436 <WAT>
A;Cross-references: GB:S66245; NID:9238706; PIDN:AAB20293.1; PID:9238707
R;Kashiwabara, S.; Baba, T.; Takada, M.; Watanabe, K.; Yano, Y.; Arai, Y.
J. Biochem. 108, 785-791, 1990
A;Title: Primary structure of mouse proacrosin deduced from the cDNA sequence and its ge
A;Reference number: JX0138; MUID:91185335; PMID:2127931

A;Accession: JX0138
A;Molecule type: mRNA
A;Residues: 4-436 <KAS>
A;Cross-references: GB:D00754; NID:9220322; PIDN:BAA00651.1; PID:9220323.
C;Comment: Acrosin is an acrosomal protease that plays an important role in the initial
e-binding activity.

C;Genetics:
A;Introns: 26/2; 95/2; 190/1; 238/3
C;Superfamily: acrosin; trypsin homology
C;Keywords: glycoprotein; hydrolase; serine proteinase; sperm
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-42/Product: acrosin light chain #status predicted <ALC>
F;43-321/Product: acrosin heavy chain #status predicted <AHC>
F;43-286/Domain: trypsin homology <TRY>
F;22,211/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;25-155,29-162,74-90,178-247,210-226,237-267/Disulfide bonds: #status predicted
F;89,143,241/Active site: His, Asp, Ser #status predicted
F;235/Binding site: substrate (Asp) #status predicted

Query Match 21.6%; Score 505; DB 2; Length 436;
Best Local Similarity 40.9%; Pred. No. 3.2e-34;
Matches 106; Conservative 37; Mismatches 90; Indels 26; Gaps 7;

QY 194 CG-----KSLKTPRWVGEEASVDSWPWQVSIQ-----DKQHVCGGSILDPHWLTA 242
Db 29 CGLRFRQNSQAGTRIVSGSQALGAWPMWVSLQIFTSHNSRRYHACGGSLLNSHWLTA 88

QY 243 HCFRKHTDVFENWK-----VRAGSKLGSPPSLA--VAKIIIIIEFNPMYPKNDIALMK 293
Db 89 HCFDNKKKVVYDWRLVFGAQEIEYGRNKPVKPEQQERYVQKIVHEKYNVVTGNDIALLK 148
QY 294 LOFFPLTFSGTVRPICLPFFDEELTPATP--LWIIWGFTKQNGKMSDILLQASVQVIDS 351
Db 149 ITPPVTGCGNFICGCLPHF-KAGPPQIPHTCYVTGWGYIKEKAPRPSFVLMEARVDLIDL 207
QY 352 TRCNADDAAYQGEVTEKMMKAGIPEGGVDTCQDSDGGPLMYQSD---QWHVVGIVSWGVC 408
Db 208 DLCNSTQWYNGRVTSTNVAGYPEGKIDTCQDSDGGPLMCRDNVDSPFVVVGITSWGVC 267
QY 409 GGPSTPGVYTKVSAYLNWI 427
Db 268 ARAKRPGVYTATWDYLDWI 286

RESULT 9
T30337
polyprotein - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 03-Feb-2003
C;Accession: T30337
R;Yang, J.C.; Lindsay, L.L.; Hedrick, J.L.
submitted to the EMBL Data Library, March 1998
A;Description: cDNA cloning of ovochymase, a chymotrypsin-like protease released from xer

A;Reference number: Z20829
A;Accession: T30337
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1524 <YAN>
A;Cross-references: EMBL:U81290; NID:92981640; PID:g2981641; PIDN:AAC24717.1
C;Superfamily: tryosin related polypeptide; trypsin homology

Query Match 21.5%; Score 504.5; DB 2; Length 1524;
Best Local Similarity 44.0%; Pred. No. 1.6e-33;
Matches 109; Conservative 38; Mismatches 86; Indels 15; Gaps 8;

QY 194 CGKSLKT----PRVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTAACHCRKH 248
Db 570 CGMAPMTPKWLPRIVVGEEASVDSWPWQVQIFFLRTFHCAGAIISQWILTAAHCIRAA 629
QY 249 TDVFENWVKVRAGS-DKLGSPPSLAIVAKIIIEFNPMYPK---DNDIALMKLOFPLTFSGTV 304
Db 630 EPSY-WTVIAGDHNRLNNESTEQIRNIKTIRIHDNVTNSETYDNDIALLYLEEPDLNDFV 688

QY 305 RPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQGEV 364
Db 689 RPYCLPEPEEVLTPASVCVVTGNTAEDGQPALG-LQQLQLPILDSIIICNT-SYSGEL 746
QY 365 TEKMMKAGIPEG-GVDTCCQDSDGGPLMYQS--DQWHVVGIVSWGVCYCGGSPSTPGVYTKVS 421
Db 747 TDHMLCAGFPSSKEKDACQDSDGGPLVCQNEKEQFSIYGLVSWGEGCGRVSKPGVYTKVR 806

QY 422 AYLNWIYN 429
Db 807 LFFTWIQN 814

RESULT 10
A37344
acrosin (EC 3.4.21.10) precursor form 1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Mar-1993 #sequence_revision 19-Mar-1993 #text_change 22-Jun-1999
C;Accession: A37344
R;Klemm, U.; Maier, W.M.; Tsaoaidou, S.; Adham, I.M.; Willison, K.; Engel, W.
Differentiation 42, 160-166, 1990
A;Title: Mouse proacrosin: cDNA sequence, primary structure and postmeiotic expessor

A;Reference number: A37344; MUID:90255839; PMID:2111255
A;Accession: A37344
A;Status: preliminary
A;Molecule type: mRNA

Db 189 PRPSPVLLAEARVELIDLCLNSTQWYNGRVMSTNVACAGYPEGKIDTCQDGGGGLMCRDN 248

QY 393 -SDQWHVVGIVSWGCGGPGSTPGVYTKVSAYLNWI 427

Db 249 ANSPFVVVGITSWGVCARAKRPGIYTATWDYLDWI 284

RESULT 13

KQHUP

plasma kallikrein (EC 3.4.21.34) precursor - human

N;Alternate names: kininogenin; plasma prekallikrein

C;Species: Homo sapiens (man)

C;Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999

C;Accession: A00921; A37939

R;Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.

Biochemistry 25, 2410-2417, 1986

A;Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four b

A;Reference number: A00921; MUID:86243359; PMID:3521732

A;Accession: A00921

A;Molecule type: mRNA

A;Residues: 1-638 <CHU>

A;Cross-references: GB:M13143; NID:G190262; PIDN:AAA60153.1; PID:G190263

R;McMullen, B.A.; Fujikawa, K.; Davie, E.W.

Biochemistry 30, 2050-2056, 1991

A;Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of

A;Reference number: A37939; MUID:91152016; PMID:1998666

A;Accession: A37939

A;Molecule type: protein

A;Residues: 20-27;40-46,'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80;103-113;131-140;141-

525;538-551;562,'X',564-567;573,'X',575-576;578-583,'X',585;592-604 <MCM>

C;Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w

are linked by one or more disulfide bonds.

C;Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a li

C;Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal

inogen and may also play a role in the renin-angiotensin system by converting prorenin i

C;Genetics:

A;Gene: GDB:KLK3

A;Cross-references: GDB:127575; OMIM:229000

A;Map position: 4q35-4q35

C;Superfamily: coagulation factor XI; trypsin homology

C;Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-638/Product: plasma kallikrein #status predicted <MAT>

F;20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>

F;20-109/Domain: apple repeat <AP1>

F;110-199/Domain: apple repeat <AP2>

F;200-289/Domain: apple repeat <AP3>

F;291-380/Domain: apple repeat <AP4>

F;391-638/Domain: plasma kallikrein light chain #status predicted <LCH>

F;391-621/Domain: trypsin homology <TRY>

F;21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,322-328,383

F;127,308,396,453,494/Binding site: carbohydrate (Asn) (covalent) #status experimental

F;318-347,340-345/Disulfide bonds: #status predicted

F;390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted

F;434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 21.0%; Score 491.5; DB 1; Length 638;

Best Local Similarity 42.2%; Pred. No. 6.8e-33;

Matches 100; Conservative 47; Mismatches 74; Indels 16; Gaps 8;

QY 202 RVVGEASVDSWPQVSIQYD---KQHVCGGSILDPHWLTAACHF--RKHTDVFNWKV 256

Db 390 RIVGNTSSGVEWPQVSLQVLTQARHLGGSLIGHQWVLTAAHCFDGLPLQDV--WRI 447

QY 257 RAG----SKLGSFPLAVAKIIIEFNPMYPKNDIALMKLPPLTFSGTVREICLPFF 312

Db 448 YSGILNLSDTKDTFQSQIKEIITHQNYKVSEGNHDIALKQAPLNYTFEQFICLPFSK 507

QY 313 DEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAYQG-EVTEKMWCA 371

Db 508 GDTSTIYNCVWTGWGFSKEK-GEIQNILQKNIPLVTNEEC--QKRYQDYKITQRMVCA 564

QY 372 GIPEGGVDTCQDGGGGLMYQ-SDQWHVVGIVSWGCGGPGSTPGVYTKVSAYLNWI 427

Db 565 GYKEGGKACKGDSGGGLVCKHNGMWRLVGTISWGEGCARREQPGVYTKVAEYMDWI 621

RESULT 14

JE0315

low-density lipoprotein receptor-related protein - mouse

C;Species: Mus musculus (house mouse)

C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Feb-2003

C;Accession: JE0315

R;Tomita, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.

J. Biochem. 124, 784-789, 1998

A;Title: A novel low-density lipoprotein receptor-related protein with type II membrane I

A;Reference number: JE0315; MUID:98429596; PMID:9756624

A;Accession: JE0315

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1113 <TOM>

A;Cross-references: DDBJ:AB013874; NID:G3869144; PIDN:BAA34371.1; PID:G3869145

C;Superfamily: mouse low-density lipoprotein receptor-related protein; LDL receptor ligai

F;337-372/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F;374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F;410-445/Domain: LDL receptor ligand-binding repeat homology <LDL3>

F;447-482/Domain: LDL receptor ligand-binding repeat homology <LDL4>

F;648-682/Domain: LDL receptor ligand-binding repeat homology <LDL5>

F;684-720/Domain: LDL receptor ligand-binding repeat homology <LDL6>

F;723-757/Domain: LDL receptor ligand-binding repeat homology <LDL7>

F;869-1097/Domain: trypsin homology <TRY>

Query Match 20.9%; Score 488.5; DB 2; Length 1113;

Best Local Similarity 33.2%; Pred. No. 2.4e-32;

Matches 130; Conservative 52; Mismatches 148; Indels 61; Gaps 17;

QY 69 IPRKQLCDGELDCPLGEDDEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDN 128

Db 736 VPRDLWCDDGWVDCSDSSDEWGCVTLSKNG-----NSSLLTVHKSKEH--HVCADG 785

QY 129 FTEALAEACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELRM-----RNSSG----- 179

Db 786 WRETLSQLACKQMGLEPSVTXLI---PGQ-----EGQWLRLYPNWNENLNGSTLQE 834

QY 180 -----PCLSGSLVSLHCLA--CGK---SLKTPRVVGGEEASVDSWPQVSIQYDKQ-HV 227

Db 835 LLVYRHSCPSRSEISLCSKQDCGRRPAARMNKRILGRTSRPGRWPWCQSLQSEPSGHI 894

QY 228 CGSILDPHWVLTAAHCFRKHDTDFVFNWVVRAGSDKL---GSFPLAVAKIIIEFNPMYP 284

Db 895 CGCVLIAKKWLTVAHCFEGREDADVWVVFGINLDPHPSGFMQTRFVKTILL--HPRYS 952

QY 285 K---DNDIALMKLPFLTFSGTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDIL 341

Db 953 RAVVDYDISVELSDDINETSIVRPVCLPSPPEXLEPDYCYITGWG---HMGNKMPFKL 1009

QY 342 LQASVQVIDSTRCNADDAY--QGEVTEKMWCAIPEGGVDTCCGDSGGPLMYQ--SDQWH 397

Db 1010 QEGEVRIIPLEQC---QSYFDMKTTITNRMICAGYESGTVDSMGDSGGPLVCERPGGQWT 1066

QY 398 VVGIVSWGCGGPGST-PGYVTKVSAYLNWI 427

Db 1067 LFGLTSWGSVCFSKVLGPGVYSNVSFYVGWI 1097

RESULT 15

S11674

acrosin (EC 3.4.21.10) precursor - human

C;Species: Homo sapiens (man)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C;Accession: S11674; S23499; S12063; A61022; S03330

R;Keime, S.; Adham, I.M.; Engel, W.

Eur. J. Biochem. 190, 195-200, 1990

A;Title: Nucleotide sequence and exon-intron organization of the human proacrosin gene.

A;Reference number: S11674; MUID:90306003; PMID:2114285

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 07:39:05 ; Search time 16.7538 Seconds
(without alignments)
1221.015 Million cell updates/sec

Title: US-09-607-745-2
Perfect score: 2342
Sequence: 1 MDPDSDQPLNSLDVKPLRKP.....VYTKVSAYLNWIYNVWKAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2337	99.8	437	1 TMS4_HUMAN	Q9nrs4 homo sapien
2	687	29.3	453	1 TMS3_MOUSE	Q8kl10 mus musculu
3	673.5	28.8	492	1 TMS2_HUMAN	O15393 homo sapien
4	660.5	28.2	454	1 TMS3_HUMAN	P57727 homo sapien
5	648.5	27.7	490	1 TMS2_MOUSE	Q9jig8 mus musculu
6	628.5	26.8	455	1 TMS5_MOUSE	Q9er04 mus musculu
7	584.5	25.0	417	1 HEPS_HUMAN	P05981 homo sapien
8	582.5	24.9	1019	1 ENTK_HUMAN	P98073 homo sapien
9	578	24.7	436	1 HEPS_MOUSE	Q35453 mus musculu
10	574	24.5	457	1 TMS5_HUMAN	Q9h3s3 homo sapien
11	571	24.4	1035	1 ENTK_BOVIN	P98072 bos taurus
12	568.5	24.3	416	1 HEPS_RAT	Q05511 rattus norv
13	568	24.3	418	1 HATT_HUMAN	O60235 homo sapien
14	552.5	23.6	1069	1 ENTK_MOUSE	P97435 mus musculu
15	552	23.6	1034	1 ENTK_PIG	P98074 sus scrofa
16	511	21.8	638	1 KAL_RAT	P14272 rattus norv
17	508	21.7	638	1 KAL_MOUSE	P26262 mus musculu
18	505	21.6	436	1 ACRO_MOUSE	P23578 mus musculu
19	503.5	21.5	324	1 TEST_MOUSE	Q9jhj7 mus musculu
20	501.5	21.4	437	1 ACRO_RAT	P29293 rattus norv
21	498	21.3	875	1 NETR_HUMAN	P56730 homo sapien
22	491.5	21.0	638	1 KAL_HUMAN	Q93952 homo sapien
23	488.5	20.9	1113	1 CORI_MOUSE	Q9z319 mus musculu
24	484	20.7	1042	1 CORI_HUMAN	Q9y5q5 homo sapien
25	483	20.6	321	1 TRYG_HUMAN	Q9nr12 homo sapien
26	482	20.6	855	1 ST14_MOUSE	P56677 mus musculu
27	481	20.5	421	1 ACRO_HUMAN	P10323 homo sapien
28	478.5	20.4	415	1 ACRO_PIG	P08001 sus scrofa
29	478	20.4	314	1 TEST_HUMAN	Q9y6m0 homo sapien
30	476.5	20.3	311	1 TRYG_MOUSE	Q9ql17 mus musculu
31	475	20.3	422	1 DES1_HUMAN	Q9ul52 homo sapien
32	472.5	20.2	317	1 BSS4_HUMAN	Q9gz44 homo sapien
33	470	20.1	761	1 NETR_MOUSE	O08762 mus musculu

34	468.5	20.0	431	1 ACRO_RABIT	P48038 oryctolagus
35	467.5	20.0	290	1 MPN_HUMAN	Q9bqr3 homo sapien
36	463	19.8	855	1 ST14_HUMAN	Q9y5y6 homo sapien
37	461.5	19.7	277	1 TRY2_ANOGA	P35036 anopheles g
38	460.5	19.7	263	1 CTRB_RAT	P07338 rattus norv
39	459.5	19.6	342	1 PSS8_RAT	Q9es87 rattus norv
40	459.5	19.6	812	1 PLMN_MOUSE	P20918 mus musculu
41	458.5	19.6	267	1 TRY7_ANOGA	P35041 anopheles g
42	458.5	19.6	790	1 PLMN_PIG	P06867 sus scrofa
43	457.5	19.5	625	1 FALL_HUMAN	P03951 homo sapien
44	452.5	19.3	342	1 PSS8_MOUSE	Q9esd1 mus musculu
45	449.5	19.2	333	1 PLMN_CANFA	P80009 canis famil

ALIGNMENTS

RESULT 1
TMS4_HUMAN STANDARD; PRT; 437 AA.
AC Q9NRS4; Q9NZA5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transmembrane protease, serine 4 (EC 3.4.21.-) (Membrane-type serine
DE protease 2) (MT-SP2).
GN TMRSS4 OR TMRSS3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancratic Carcinoma;
RX MEDLINE=20283276; PubMed=10825129;
RA Wallrapp C., Haehnel S., Mueller-Pillasch F., Burghardt B.,
RA Iwamura T., Ruthenbuerger M., Lerch M.M., Adler G., Gress T.M.;
RT "A novel transmembrane serine protease (TMRSS3) overexpressed in
RT pancreatic cancer.";
RL Cancer Res. 60:2602-2606(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Smeekens S.S., Lorimer D.D., Wang E., Hou J., Linnevers C.;
RT "MT-SP2, a novel type II membrane serine protease expressed in
RT trachea, colon, and small intestine: identification, cloning, and
RT chromosomal localization.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: MAY PLAY A ROLE IN TUMOR PROGRESSION, METASTASIS
FORMATION AND TUMOR INVASION.

-!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
-!- TISSUE SPECIFICITY: HIGH LEVELS IN PANCREATIC, GASTRIC, COLORECTAL AND AMPULLARY CANCER. VERY WEAK EXPRESSION IN NORMAL GASTROINTESTINAL AND UROGENITAL TRACT.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
-!- SIMILARITY: Contains 1 SRCR domain.
-!- SIMILARITY: Contains 1 LDL-receptor class A domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF179224; AAF74526.1; -.
EMBL; AF216312; AAF31436.1; -.
EMBL; BC011703; AAH11703.1; -.
HSSP; P00763; LDPO.
MEROPS; S01.034; -.
Genew; HGNC:11878; TMPRSS4.
MIM; 606565; -.
GO; GO:0016021; C:integral to membrane; NAS.
GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
InterPro; IPR001314; Chymotrypsin.
InterPro; IPR002172; LDL receptor A.
InterPro; IPR001254; Ser_protease_Try.
InterPro; IPR001190; Srcr_receptor.
Pfam; PF00057; ldl_recept_a; 1.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SM00192; LDLa; 1.
SMART; SM00202; SR; 1.
SMART; SM00020; Tryp_Spc; 1.
PROSITE; PS01209; LDLRA_1; FALSE_NEG.
PROSITE; PS50068; LDLRA_2; FALSE_NEG.
PROSITE; PS00420; SRCR_1; FALSE_NEG.
PROSITE; PS50287; SRCR_2; 1.
PROSITE; PS50240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SER; 1.
Hydrolase; Serine protease; Transmembrane; Signal-anchor.
CYTOPLASMIC (POTENTIAL).
SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
LDL-RECEPTOR CLASS A.
SRCR.
SERINE PROTEASE.
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
CLEAVAGE (POTENTIAL).
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
MLQDPDSQPLNSLDVKPLRKPRIPMETFRK -> MSNPCA
NPVSPWRPSES (IN REF. 2).
SEQUENCE 437 AA; 48204 MW; 351B2FD4A8657B12 CRC64;
Query Match 99.8%; Score 2337; DB 1; Length 437;
Best Local Similarity 100.0%; Pred.No. 2.1e-195;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYFL 61

Db 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYFL 63
QY 62 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAPAVRLSKDRSTLQVLDATGNW 121
Db 64 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAPAVRLSKDRSTLQVLDATGNW 123
QY 122 FSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLVDVVEITENSQELMRNSSGPC 181
Db 124 FSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLVDVVEITENSQELMRNSSGPC 183
QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGGSILDPHWILTA 241
Db 184 LSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGGSILDPHWILTA 243
QY 242 AHCERKHTDVFNWVKVRAGSKLGSFPLAVAKIIIEFNPMYPKNDIALMKLOFPLTFS 301
Db 244 AHCERKHTDVFNWVKVRAGSKLGSFPLAVAKIIIEFNPMYPKNDIALMKLOFPLTFS 303
QY 302 GTVRPICLPFFDEELTPATPLWIIGWFTKQGGKMSDILLQASVQVIDSTRCNADDAYQ 361
Db 304 GTVRPICLPFFDEELTPATPLWIIGWFTKQGGKMSDILLQASVQVIDSTRCNADDAYQ 363
QY 362 GEVTEKMMCAGIPEGGVDTCQGSGLPLMYQSDQHVHVGIVSWGCGGPGSTPGVYTKVS 421
Db 364 GEVTEKMMCAGIPEGGVDTCQGSGLPLMYQSDQHVHVGIVSWGCGGPGSTPGVYTKVS 423
QY 422 AYLNWIYVNWKAEL 435
Db 424 AYLNWIYVNWKAEL 437
RESULT 2
TMS3_MOUSE
ID TMS3_MOUSE STANDARD; PRT; 453 AA.
AC Q8K1T0; Q8VDE0;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transmembrane protease, serine 3 (EC 3.4.21.-).
GN TMPRSS3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND FUNCTION IN ENAC
RP CLEAVAGE.
RX PubMed=12393794;
RA Guipponi M., Vuagniaux G., Wattenhofer M., Shibuya K., Vazquez M.,
RA Dougherty L., Scamuffa N., Guida E., Okui M., Rossier C., Hancock M.,
RA Buchet K., Reymond A., Hummler E., Marzella P.L., Kudoh J.,
RA Shimizu N., Scott H.S., Antonarakis S.E., Rossier B.C.;
RT "The transmembrane serine protease (TMPRSS3) mutated in deafness
RT DFNB8/10 activates the epithelial sodium channel (ENaC) in vitro.";
RL Hum. Mol. Genet. 11:2829-2836(2002).
CC -!- FUNCTION: Probable protease. Seems to be capable of activating
CC ENaC.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
CC reticulum.
CC -!- TISSUE SPECIFICITY: Expressed in the spiral ganglion, the cells
CC supporting the organ of Corti and the stria vascularis.
CC -!- PTM: Undergoes autolytic activation.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: Contains 1 SRCR domain.
CC -!- SIMILARITY: Contains 1 LDL-receptor class A domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

or send an email to license@isb-sib.ch).

CC	CC	DR	EMBL; AJ429216; CAD22137.1; -	
DR	EMBL; AJ300738; CAC83350.1; -			
DR	HSSP; P00761; 1AN1.			
DR	MGI; MGI:2155445; Tmprs3.			
DR	InterPro; IPR001314; Chymotrypsin.			
DR	InterPro; IPR002172; LDL_receptor_A.			
DR	InterPro; IPR001254; Ser_protease_Try.			
DR	InterPro; IPR001190; Srcr_receptor.			
DR	Pfam; PF00057; ldl_recept_a; 1.			
DR	Pfam; PF00089; trypsin; 1.			
DR	PRINTS; PR00722; CHYMOTRYPSIN.			
DR	SMART; SMO0192; LDLa; 1.			
DR	SMART; SMO0202; SR; 1.			
DR	SMART; SMO0020; Tryp_Spc; 1.			
DR	PROSITE; PS01209; LDLRA_1; 1.			
DR	PROSITE; PS50068; LDLRA_2; 1.			
DR	PROSITE; PS50287; SRCR_2; 1.			
DR	PROSITE; PS50240; TRYPSIN_DOM; 1.			
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.			
DR	PROSITE; PS00135; TRYPSIN_SER; 1.			
KW	Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;			
KW	Endoplasmic reticulum.			
FT	DOMAIN 1 48			
FT	TRANSMEM 49 69			
FT				
FT	DOMAIN 70 453			
FT	DOMAIN 72 108			
FT	DOMAIN 104 205			
FT	DOMAIN 217 448			
FT	ACT_SITE 257 257			
FT	ACT_SITE 304 304			
FT	ACT_SITE 400 400			
FT	SITE 216 217			
FT	DISULFID 73 85			
FT	DISULFID 79 98			
FT	DISULFID 92 107			
FT	DISULFID 129 194			
FT	DISULFID 142 204			
FT	DISULFID 207 324			
FT	DISULFID 242 258			
FT	DISULFID 338 406			
FT	DISULFID 369 385			
FT	DISULFID 396 424			
FT	CARBOHYD 221 221			
FT	CONFLICT 117 117			
FT	CONFLICT 246 246			
SQ	SEQUENCE 453 AA; 49491 MW; 1ABCBF10AF6B1EF6 CRC64;			
Query Match 29.3%; Score 687; DB 1; Length 453;				
Best Local Similarity 36.9%; Pred. No. 5e-52;				
Matches 161; Conservative 72; Mismatches 153; Indels 50; Gaps 14;				
QY	22	IPMETFRKVGIPPIIITALLSLASIIIVVLIKVILD---	KYFLCGQLHFIIPRKQLCDGE 78	
Db	44	LPLKFPPIIVIGIILALALA-----	IGLGHFDCSGKY--RCHSFKCIELTARCDGV 95	
QY	79	LDCPLGEDEEHCVKSPFEGPAVAVRLSKORSTLQVLDSATGNWFSACFDNFTEALAEATAC 138		
Db	96	SDCKNAEYRC-----	VRVSGQRAALQVFTAA--WRTMCSDDKSHYAKIAC 142	
QY	139	RQMGYSS-----	KPTFRAVEIGPDQLDVEITENSQELMRNSSGPCLSGSL 186	
Db	143	AQLGFPSYVSSDHLRVDALBEEQFGDFVSINHLSSDDKVLTALHSHVYMREG---	CTSGHV 199	
QY	187	VSLHCLACG-KSLKTRPVVGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLTAACHF 245		
Db	200	VTLKCSACGTRTGYSRIVGGNMSSLTQWPQVSLQFGYHLCGGSVITPLWIVTAACHV 259		
QY	246	RKHTDVFNWKRAGSKL--	GSFPSLAVAKIIIIIEFNPMY-PK--DNDIALMKLQFPLTF 300	
Db	260	YDLYHPKSWTVQVGLVSLMDSVPVSHLVEKII--	YHSKYKPKRLGNDIALMKLSEPLTF 316	

QY	301	SCTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAY 360		
Db	317	DETIQPICLPNSENFPDGLKLCWTSGWGAT-EDGGDASPLNHAAPLISNKNICNHRDVG 375		
QY	361	QGEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQ-WHVVGIVSWGCGGPGSTPGVYTK 419		
Db	376	GGIISPSMLCAGYLKGGVDSQCGDSGGPLVCQERRLWKLVGATSGFGIGCAEVNKPVGYYTR 435		
QY	420	VSAYLNWIYNVWKAEHL 435		
Db	436	ITSFLDWIHEQLERDL 451		

RESULT 3

TMS2_HUMAN				
ID	TMS2_HUMAN	STANDARD;	PRT;	492 AA.
AC	O15393; Q9BXX1;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	16-OCT-2001 (Rel. 40, last sequence update)			
DT	28-FEB-2003 (Rel. 41, last annotation update)			
DE	Transmembrane protease, serine 2 precursor (EC 3.4.21.-).			
GN	TPRSS2 OR PRSS10.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97468144; PubMed=9325052;			
RA	Paoloni-Giacobino A., Chen H., Peitsch M.C., Rossier C.,			
RA	Antonarakis S.E.;			
RT	"Cloning of the TPRS2 gene, which encodes a novel serine protease			
RT	with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3.";			
RL	Genomics 44:309-320(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21309069; PubMed=11414763;			
RA	Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;			
RT	"Mutation analyses of 268 candidate genes in human tumor cell lines.";			
RL	Genomics 74:352-364(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A., AND MUTAGENESIS.			
RX	MEDLINE=21139112; PubMed=11245484;			
RA	Afar D.E.H., Vivanco I., Hubert R.S., Kuo J., Chen E., Saffran D.C.,			
RA	Raitano A.B., Jakobovits A.;			
RT	"Catalytic cleavage of the androgen-regulated TPRS2 protease results			
RT	in its secretion by prostate and prostate cancer epithelia.";			
RL	Cancer Res. 61:1686-1692(2001).			
RN	[4]			
RP	TISSUE SPECIFICITY.			
RX	MEDLINE=21104370; PubMed=11169526;			
RA	Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viikko P.T.;			
RT	"Expression of transmembrane serine protease TPRS2 in mouse and			
RL	human tissues.";			
CC	J. Pathol. 193:134-140(2001).			
CC	-I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY			
CC	CLEAVAGE AND SECRETED.			
CC	-I- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE. ALSO			
CC	EXPRESSED IN PROSTATE, COLON, STOMACH, AND SALIVARY GLAND.			
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.			
CC	-I- SIMILARITY: Contains 1 SRCR domain.			
CC	-I- SIMILARITY: Contains 1 LDL-receptor class A domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U75329; AAC51784.1; -.			

families with non-syndromic autosomal recessive deafness."; Hum. Mutat. 18:101-108(2001).

[5]

VARIANTS DFBN8/DFNB10 TRP-109; PHE-194 AND ARG-407, AND VARIANTS ILE-53; SER-111 AND VAL-253.

RP MEDLINE=21317610; PubMed=11424922;

RA Ben-Yosef T., Wattenhofer M., Riazuddin S., Ahmed Z.M., Scott H.S., Kudoh J., Shibuya K., Antonarakis S.E., Bonne-Tamir B., Radhakrishna U., Naz S., Ahmed Z., Riazuddin S., Pandya A., Nance W.E., Wilcox E.R., Friedman T.B., Morell R.J.;

RA "Novel mutations of TMPS3 in four DFBN8/B10 families segregating congenital autosomal recessive deafness."; J. Med. Genet. 38:396-400(2001).

[6]

VARIANT DFBN8/DFNB10 GLY-103, AND VARIANTS ASN-173 AND THR-426. MEDLINE=21904597; PubMed=11907649;

RA Wattenhofer M., Di Iorio V., Rabionet R., Dougherty L., Pampanos A., Schwede T., Montserrat-Sentis B., Arbones L., Iliades T., Pasquadi bisceglie A., D'Amelio M., Alwan S., Rossier C., Dahl H.-H.M., Petersen M.B., Estivill X., Gasparini P., Scott H.S., Antonarakis S.E.;

RA "Mutations in the TMPS3 gene are a rare cause of childhood nonsyndromic deafness in Caucasian patients."; J. Mol. Med. 80:124-131(2002).

CC -!- FUNCTION: Probable protease. Seems to be capable of activating ENaC.

CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic reticulum.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=4;

CC Name=A;

CC IsoId=P57727-1; Sequence=Displayed;

CC Name=B; Synonyms=C;

CC IsoId=P57727-2; Sequence=VSP_005391;

CC Name=D;

CC IsoId=P57727-3; Sequence=VSP_005392;

CC Name=T; Synonyms=truncated, TAG-12V;

CC IsoId=P57727-4; Sequence=VSP_005393, VSP_005394;

CC -!- TISSUE SPECIFICITY: Expressed in many tissues including fetal cochlea. Isoform T is found at increased levels in some carcinomas.

CC -!- PTM: Undergoes autophosphorylation.

CC -!- DISEASE: Defects in TMPS3 are a cause of childhood-onset autosomal recessive neurosensory deafness 8 (DFNB8) [MIM:601072].

CC -!- DISEASE: Defects in TMPS3 are a cause of congenital autosomal recessive neurosensory deafness 10 (DFNB10) [MIM:605316].

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

CC -!- SIMILARITY: Contains 1 SRCR domain.

CC -!- SIMILARITY: Contains 1 LDL-receptor class A domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AF201380; AAC37012.1; -.

CC EMBL; AB038157; BAB20077.1; -.

CC EMBL; AB038158; BAB20078.1; -.

CC EMBL; AB038159; BAB20079.1; -.

CC EMBL; AB038160; BAB20080.1; -.

CC HSSP; P00763; 1DPO.

CC MEROPS; S01.079; -.

CC Genew; HGNC:11877; TMPS3.

CC MIM; 605511; -.

CC MIM; 601072; -.

CC MIM; 605316; -.

CC GO; GO:0016021; C:integral to membrane; NAS.

CC GO; GO:0006508; P:proteolysis and peptidolysis; NAS.

CC InterPro; IPR001314; Chymotrypsin.

CC InterPro; IPR002172; LDL_receptor_A.

DR InterPro; IPR001254; Ser_protease_Try.

DR InterPro; IPR001190; Srcr_receptor.

DR Pfam; PF00057; ldl_recept_a; 1.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00192; LDLa; 1.

DR SMART; SM00202; SR; 1.

DR SMART; SM00020; Tryp_SPC; 1.

DR PROSITE; PS01209; LDLRA_1; 1.

DR PROSITE; PS00068; LDLRA_2; 1.

DR PROSITE; PS00420; SRCR_1; FALSE_NEG.

DR PROSITE; PS0287; SRCR_2; 1.

DR PROSITE; PS0240; TRYPSIN_DOM; 1.

DR PROSITE; PS00134; TRYPSIN_HIS; 1.

DR PROSITE; PS00135; TRYPSIN_SER; 1.

DR Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;

KW Endoplasmic reticulum; Deafness; Alternative splicing;

KW Disease mutation; Polymorphism.

FT DOMAIN 1 48 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 49 69 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).

FT DOMAIN 70 454 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 72 108 LDL-RECEPTOR CLASS A.

FT DOMAIN 109 205 SRCR.

FT DOMAIN 217 454 SERINE PROTEASE.

FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 304 304 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 401 401 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT SITE 216 217 CLEAVAGE (POTENTIAL).

FT DISULFID 73 85 BY SIMILARITY.

FT DISULFID 79 98 BY SIMILARITY.

FT DISULFID 92 107 BY SIMILARITY.

FT DISULFID 129 194 BY SIMILARITY.

FT DISULFID 142 204 BY SIMILARITY.

FT DISULFID 207 324 BY SIMILARITY.

FT DISULFID 242 258 BY SIMILARITY.

FT DISULFID 338 407 BY SIMILARITY.

FT DISULFID 370 386 BY SIMILARITY.

FT DISULFID 397 425 BY SIMILARITY.

FT CARBOHYD 221 221 N-LINKED (GLCNAC...) (POTENTIAL).

FT VARSPLIC 1 127 Missing (in isoform B).

FT VARSPLIC 318 454 /FTid=VSP_005391.

EMIQVCLPNSEENFPDGKVCWTSGWATEDGAGDASPVLN

HAAPVLSNKNICNHRDVGGIISPSMLCAGYLTGGVDSQCG

DSGGLVCQERRLLKLVGATSGIGCAEVNKPQVYTRVTSF

LDWIHEQMERDLKT -> GTSGSLCGSAALPLFQEDLQLLI

EAFI (in isoform D).

/FTid=VSP_005392.

DLXLPKSWTQVGLVSLDNPAPSHLVEKIVYH -> RIVA

PRERADRRGRKLLCWRKPTKMGPRPSHS (in isoform T).

/FTid=VSP_005393.

Missing (in isoform T).

/FTid=VSP_005394.

V -> I.

/FTid=VAR_010781.

D -> G (in DFBN8/DFNB10).

/FTid=VAR_013490.

R -> W (in DFBN8/DFNB10).

/FTid=VAR_013491.

G -> S.

/FTid=VAR_013492.

D -> N.

/FTid=VAR_013493.

C -> F (in DFBN8/DFNB10).

/FTid=VAR_013494.

W -> C (in DFBN8/DFNB10).

/FTid=VAR_011678.

I -> V (in dBSNP:2839500).

/FTid=VAR_013101.

P -> L (in DFBN8/DFNB10).

/FTid=VAR_011679.

C -> R (in DFBN8/DFNB10).

b 46 -----SDQE-----PLYPVQSSADARLWVFDKTEGTWRLLCSSRSNARVAGLSC 90
y 139 RQMGYSKPTFRAVEIGPDQDLVDVEITEN-----SQELMRNSSGPGCLS 183
b 91 EEMGF-----LRAL-----THSELVDRTAGANGTSFFCVDEGRLPHTQRLLEVISVCDPCR 142
y 184 GSLVSLHCLACG-KSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLTA 242
b 143 GRFLAAICQDCGRKLPVDRIVGGRDTSLGRWPQVSLRYDGAHLGCGSLLSGDWLTA 202
y 243 HCFRKHTDVFN-WKVRACSKLGSFPSPSLAVAKIILIEFNPMP-----KNDIALMK 293
b 203 HCFPERNVLRSRVRVAGAVAQASPHGLQLGVQAVVYHGGLPFRDPENSENSNDIALVH 262
y 294 LQFPLTFSGTVRPICLPFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTR 353
b 263 LSSPLPLTEYIQPVCLPAAGQALVDGKICTVTGWNT-QYVGQAGVLQEARVPIISNDV 321
y 354 CNADDAYQGEVTEKMMKAGIPGEGVDTCCQSDSGGLMYQ-----SDQWHVVGVISWGYG 408
b 322 CNGADFYGNQIKPKMFCAGYPEGGIDACQDSSGGPFVCEDSISRTPRWRLCGIVSWGTC 381
y 409 GGPSTPGVYTKVSAYLWNIYNWK 432
b 382 ALAQKPGVYTKVSDFREWIFQA 405

RESULT 8

NTK_HUMAN D ENT_K_HUMAN STANDARD; PRT; 1019 AA.
C P98073;
T 01-FEB-1996 (Rel. 33, Created)
T 01-FEB-1996 (Rel. 33, Last sequence update)
T 15-SEP-2003 (Rel. 42, Last annotation update)
E Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
S PRSS7 OR ENTK.
S Homo sapiens (Human).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
X NCBI_TaxID=9606;
X [1]
P SEQUENCE FROM N.A.
C TISSUE=Duoenum;
X MEDLINE=95234679; PubMed=7718557;
X Kitamoto Y., Veile R.A., Donis-Keller H., Sadler J.E.;
A "cDNA sequence and chromosomal localization of human enterokinase,
T the proteolytic activator of trypsinogen.";
L Biochemistry 34:4562-4568(1995).
[2]
R SEQUENCE FROM N.A., AND DISEASE.
X MEDLINE=21606074; PubMed=11719902;
R Holzinger A., Maier E.M., Buck C., Mayerhofer P.U., Kappler M.,
R Haworth J.C., Moroz S.P., Hadorn H.-B., Sadler J.E., Roscher A.A.;
A "Mutations in the proenteropeptidase gene are the molecular cause of
T congenital enteropeptidase deficiency.";
L Am. J. Hum. Genet. 70:20-25(2002).
[3]
R SEQUENCE FROM N.A.
X MEDLINE=20289799; PubMed=10830953;
R Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
R Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
R Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
R Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
R Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
R Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
R Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
R Minoshima S., Shimizu N., Nordsiek G., Hornisiek K., Brandt P.,
R Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
R Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
R Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
R Lehrach H., Reinhardt R., Yaspo M.-L.;
A "The DNA sequence of human chromosome 21.";

RL Nature 405:311-319(2000).
RN [4]
RP SEQUENCE OF 749-1019 FROM N.A.
RC TISSUE=Duoenum;
RX MEDLINE=94329561; PubMed=8052624;
RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
CC -!- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in
CC trypsinogen.
CC -!- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -!- DISEASE: Defects in PRSS7 are a cause of enterokinase deficiency
CC [MIM:226200]; a life-threatening intestinal malabsorption disorder
CC characterized by diarrhea and failure to thrive.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC -!- SIMILARITY: Contains 1 SRCR domain.
CC -!- SIMILARITY: Contains 1 MAM domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U09860; AAC50138.1; -.
DR EMBL; Y19124; CAB65555.1; -.
DR EMBL; Y19125; CAB65555.1; JOINED.
DR EMBL; Y19126; CAB65555.1; JOINED.
DR EMBL; Y19127; CAB65555.1; JOINED.
DR EMBL; Y19128; CAB65555.1; JOINED.
DR EMBL; Y19129; CAB65555.1; JOINED.
DR EMBL; Y19130; CAB65555.1; JOINED.
DR EMBL; Y19131; CAB65555.1; JOINED.
DR EMBL; Y19132; CAB65555.1; JOINED.
DR EMBL; Y19133; CAB65555.1; JOINED.
DR EMBL; Y19134; CAB65555.1; JOINED.
DR EMBL; Y19135; CAB65555.1; JOINED.
DR EMBL; Y19136; CAB65555.1; JOINED.
DR EMBL; Y19137; CAB65555.1; JOINED.
DR EMBL; Y19138; CAB65555.1; JOINED.
DR EMBL; Y19139; CAB65555.1; JOINED.
DR EMBL; Y19140; CAB65555.1; JOINED.
DR EMBL; Y19141; CAB65555.1; JOINED.
DR EMBL; Y19142; CAB65555.1; JOINED.
DR EMBL; Y19143; CAB65555.1; JOINED.
DR EMBL; AL163218; CAB90392.1; -.
DR EMBL; AL163217; CAB90389.1; -.
DR PIR; A56318; A56318.
DR HSSP; P00763; 1DPO.
DR MEROPS; S01.156; -.
DR Genew; HGNC:9490; PRSS7.
DR MIM; 606635; -.
DR GO; 0005903; C:brush border; TAS.
DR GO; 0005903; C:brush border; TAS.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000859; CUB domain.
DR InterPro; IPR002172; LDL_receptor_A.

DR InterPro; IPR000998; MAM domain.
DR InterPro; IPR000082; SEA domain.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; ldl_recept_a; 2.
DR Pfam; PF00629; MAM; 1.
DR Pfam; PF01390; SEA; 1.
DR Pfam; PF00530; SRCR; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00192; LDLA; 2.
DR SMART; SM00137; MAM; 1.
DR SMART; SM00200; SEA; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS00068; LDLRA_2; 2.
DR PROSITE; PS00740; MAM_1; 1.
DR PROSITE; PS00060; MAM_2; 1.
DR PROSITE; PS00024; SEA; 1.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Signal-anchor; Glycoprotein; Myristate; Hydrolase;
KW Serine protease; Zymogen; Transmembrane; Repeat.
FT CHAIN 1 784 NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT CHAIN 785 1019 CATALYTIC CHAIN (LIGHT CHAIN).
FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 48 1019 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 52 169 SEA.
FT DOMAIN 182 223 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 225 334 CUB 1.
FT DOMAIN 342 504 MAM.
FT DOMAIN 524 634 CUB 2.
FT DOMAIN 641 679 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 678 771 SRCR.
FT DOMAIN 785 1019 SERINE PROTEASE.
FT ACT_SITE 825 825 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 876 876 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 971 971 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT LIPID 2 2 MYRISTATE (POTENTIAL).
FT DISULFID 184 197 BY SIMILARITY.
FT DISULFID 191 210 BY SIMILARITY.
FT DISULFID 204 221 BY SIMILARITY.
FT DISULFID 643 655 BY SIMILARITY.
FT DISULFID 650 668 BY SIMILARITY.
FT DISULFID 662 677 BY SIMILARITY.
FT DISULFID 772 896 INTERCHAIN (BY SIMILARITY).
FT DISULFID 810 826 BY SIMILARITY.
FT DISULFID 910 977 BY SIMILARITY.
FT DISULFID 941 956 BY SIMILARITY.
FT DISULFID 967 995 BY SIMILARITY.
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 335 335 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 706 706 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 134 134 Q -> E (IN REF. 3).
FT CONFLICT 732 732 S -> P (IN REF. 3).
FT CONFLICT 754 754 SQCLQDLSLRQC�NHKS -> RRNAKNEIDALSPIILIA
(IN REF. 3).
SQ SEQUENCE 1019 AA; 112923 MW; B6AAA245F6D4A563 CRC64;
Query Match 24.9%; Score 582.5; DB 1; Length 1019;
Best Local Similarity 33.8%; Pred. No. 1.7e-42;
Matches 135; Conservative 62; Mismatches 132; Indels 71; Gaps 15;
QY 67 HF-----IPRKQLCDGELDPLGEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATG 119
DB 647 HFQCKNGECVPLVNLCDGHLHCEGDSDEADCVRF-----NGTTN 686
QY 120 N-----WFSACFDNPTTEALAEATACROMGY-----SSKPTFRAVEIGPDQDLDVVE 164
DB 687 NGLVRFRIQSIWHTACAEWNTQISNDVCQLLGLSGNSKPIF-STDCGPFVVKLNTAP 745
QY 165 ----ITENSQELMRNRSNGPCISGSLVSLHC--LACKSLK---TPRVVCGEEASVDSW 214
DB 746 DGHLLTPSQO-----CLQDSLIRLQCNHKSCGKKLAAQDITPKIVGGSNAKEGAW 796
QY 215 PWQVSIQDKQHVCGGSILDPHWLTAHHC-FRKHTDVFNWVKVRAGSKLGSPPSLAVAK 273
DB 797 PWVVGLYGGRLLCGASLVSSDWLSAAHCVYGRNLEPFSKWTAILGLHMKSNLTSPQTV 856
QY 274 III--IEFNMY--PKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPATPLWIIGWG 328
DB 857 RLIDEIVNPHNRRKONDIAHMHLEFKVNYTDYQICLPEENQVFPGRNCISAGWG 916
QY 329 FTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMCAGIPEGGVDTCCQDSSGGP 388
DB 917 -TVVYQGTANILQEADVPLLSNERCQ-QQMPEYNITENMICAGYEEGGIDSCQDSSGGP 974
QY 389 LMYQ-SDQWHVVGIVSWGVCGGSPSTPGVYTKVSAYLNWI 427
DB 975 LMCQENRWFLAGVTSFGYKCALPNRPVGVYARVSRFTWI 1014

RESULT 9
HEPS_MOUSE
ID HEPS_MOUSE STANDARD; PRT; 436 AA.
AC O35453; Q9CW97;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Serine protease hepsin (EC 3.4.21.-).
GN HPN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Liver;
RX MEDLINE=98058912; PubMed=9395459;
RA Vu T.-K.H., Liu R.W., Haakma C., Tomasek J.J., Howard E.W.;
RT "Identification and cloning of the membrane-associated serine
protease, hepsin, from mouse preimplantation embryos.";
RL J. Biol. Chem. 272:31315-31320(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=99339944; PubMed=10411637;
RA Kawamura S., Kurachi S., Deyashiki Y., Kurachi K.;
RT "Complete nucleotide sequence, origin of isoform and functional
characterization of the mouse hepsin gene.";
RL Eur. J. Biochem. 262:755-764(1999).

```

RN SEQUENCE FROM N.A. (ISOFORM 1).
RP STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: Plays an essential role in cell growth and maintenance
CC of cell morphology.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=1a;
CC IsoId=O35453-1; Sequence=Displayed;
CC Note=Minor isoform;
CC Name=2; Synonyms=2a;
CC IsoId=O35453-2; Sequence=VSP_007232;
CC Note=Major isoform;
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to
CC frameshifts in positions 155, 191 and 233.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF030065; AAB84221.1; .
CC EMBL; AK002694; BAB22289.2; ALT_FRAME.
CC HSSP; P00763; IDPO.
CC MEROPS; S01.224; -.
CC MGD; MGI:1196620; Hpn.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR001254; Ser_protease_Try.
CC InterPro; IPR001190; Srrc_receptor.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00202; SR; 1.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PS50240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Hydrolase; Serine protease; Transmembrane; Signal-anchor;
CC Alternative splicing.
CC CHAIN 1 181 SERINE PROTEASE HEPsin, NON-CATALYTIC
CC CHAIN 182 436 CHAIN (POTENTIAL).
CC FT FT SERINE PROTEASE HEPsin, CATALYTIC CHAIN
CC FT FT (POTENTIAL).
CC DOMAIN 21 36 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 37 63 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC FT FT (POTENTIAL).
CC DOMAIN 64 436 EXTRACELLULAR (POTENTIAL).
CC FT FT SERINE PROTEASE.

```

```

FT ACT_SITE 222 222 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 276 276 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 372 372 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 172 296 INTERCHAIN (BY SIMILARITY).
FT DISULFID 207 223 BY SIMILARITY.
FT DISULFID 341 357 BY SIMILARITY.
FT DISULFID 368 400 BY SIMILARITY.
FT CARBOHYD 131 131 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT VARSPLIC 25 44 Missing (in isoform 2).
FT CONFLICT 85 85 /FTId=VSP_007232.
FT CONFLICT 204 204 L -> F (IN REF. 2 AND 3).
FT CONFLICT 214 214 T -> Y (IN REF. 3).
FT CONFLICT 228 229 G -> R (IN REF. 3).
FT CONFLICT 264 264 NR -> ET (IN REF. 3).
FT CONFLICT 281 281 P -> L (IN REF. 3).
FT CONFLICT 281 281 H -> N (IN REF. 3).
SQ SEQUENCE 436 AA; 46787 MW; 4A0993148C620BD0 CRC64;

Query Match 24.7%; Score 578; DB 1; Length 436;
Best Local Similarity 31.1%; Pred. No. 1.4e-42;
Matches 136; Conservative 62; Mismatches 150; Indels 90; Gaps 10;

QY 30 VGPIIIALLSLASIIIVVLIKVILDKYFLCGQLHFIPRKQLCDGELDCPLGEDEEH 89
Db 42 VGTLLFLTGTGGAASWAIVTILQ-----SDQE- 68
QY 90 CVKSPFEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDNFTEALAEATACRMGY----- 143
Db 69 -----PLYQQLSPGDSRLAVLDKTEGTWRLLCSSRSNARVAGLCEMGFLRALAH 120
QY 144 -----SSKPTFRAVEIG---PDQDLDVVEITENSQELRMENSSGPGCLSGSLVSL 189
Db 121 SELDVRTAGANGTSGFFCVDDEGLPLAQLLDVISVCD-----CPRGRFLTA 167
QY 190 HCLACG-KSLKTPRVVGEEASVDSWPWQVSIQYDKQHVCGGSLDPHWVLTAAHCFRKH 248
Db 168 TCQDCGRRKLPVDRIVGGQDSSLGRWPWQVSLRYDGTGLCGGSLSGDWVLTAAHCFPER 227
QY 249 TDVFN-WKVRACSKLGSFPSSLAVAKIIIEFNPMPY-----KNDIALMKLQFPLT 299
Db 228 NRVLRSRVRVAGAVARTSPHAVQLGVQAVIYHGGYLPFRDPTIDENSNDIALVHLSLSP 287
QY 300 FSGTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDA 359
Db 288 LLEYIQPVCPLPAAGQALVDGKVCVTVGWNT-QFYGQQAAMVLEARVPIISNEVCNSPDF 346
QY 360 YQGEVTEKMMCAGIPEGGVDTCCQSDSGPLMYQ-----SDQHVVGIVSWGCGGPSTP 414
Db 347 YGNQIKPKMFCAGYPEGGIDACQSDSGPFVCEDSISGTSRWRLCGIVSWGTCALARKP 406
QY 415 GVTYKVSAYLNWIYNVWK 432
Db 407 GVTYKVTDFREWIFKAIK 424

RESULT 10
TMS5_HUMAN
ID TMS5_HUMAN STANDARD; PRT; 457 AA.
AC Q9H353;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
GN TMRSS5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX PubMed=11741986;
RA Yamaguchi N., Okui A., Yamada T., Nakazato H., Mitsui S.;

```



```

-!- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
-!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
    CLEAVED BY A TRYPSIN-LIKE PROTEASE.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
-!- SIMILARITY: Contains 2 LDL-receptor class A domains.
-!- SIMILARITY: Contains 2 CUB domains.
-!- SIMILARITY: Contains 1 SEA domain.
-!- SIMILARITY: Contains 1 SRCR domain.
-!- SIMILARITY: Contains 1 MAM domain.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

EMBL; U09859; AAB40026.1; -.
EMBL; L19663; AAA16035.1; -.
PIR; A43090; A43090.
PDB; 1EKB; 14-OCT-99.
MEROPS; S01.156; -.
InterPro; IPR001314; Chymotrypsin.
InterPro; IPR000859; CUB domain.
InterPro; IPR002172; LDL_receptor_A.
InterPro; IPR000998; MAM_domain.
InterPro; IPR000082; SEA_domain.
InterPro; IPR001254; Ser_protease_Try.
InterPro; IPR001190; Srcr_receptor.
Pfam; PF00431; CUB; 2.
Pfam; PF00057; ldl_recept_a; 2.
Pfam; PF00629; MAM; 1.
Pfam; PF01390; SEA; 1.
Pfam; PF00530; SRCR; 1.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
PRINTS; PR00261; LDLRECEPTOR.
SMART; SM00042; CUB; 2.
SMART; SM00192; LDLa; 2.
SMART; SM00137; MAM; 1.
SMART; SM00200; SEA; 1.
SMART; SM00202; SR; 1.
SMART; SM00020; Tryp_SPC; 1.
PROSITE; PS01180; CUB; 2.
PROSITE; PS01209; LDLRA_1; 2.
PROSITE; PS50068; LDLRA_2; 2.
PROSITE; PS00740; MAM_1; 1.
PROSITE; PS50060; MAM_2; 1.
PROSITE; PS50024; SEA; 1.
PROSITE; PS00420; SRCR_1; FALSE_NEG.
PROSITE; PS50287; SRCR_2; 1.
PROSITE; PS50240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SER; 1.
Signal-anchor; Glycoprotein; Myristate; Hydrolase; Serine protease;
Zymogen; Transmembrane; Repeat; Alternative splicing; 3D-structure.
CHAIN 1 800 NON-CATALYTIC CHAIN (HEAVY CHAIN).
CHAIN 801 1035 CATALYTIC CHAIN (LIGHT CHAIN).
DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
DOMAIN 48 1035 EXTRACELLULAR (POTENTIAL).
DOMAIN 54 169 SEA.
DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.
DOMAIN 240 350 CUB 1.
DOMAIN 358 520 MAM.
DOMAIN 540 650 CUB 2.
DOMAIN 657 695 LDL-RECEPTOR CLASS A 2.
DOMAIN 694 787 SRCR.
DOMAIN 801 1035 SERINE PROTEASE.
ACT_SITE 841 841 CHARGE RELAY SYSTEM (BY SIMILARITY).
ACT_SITE 892 892 CHARGE RELAY SYSTEM (BY SIMILARITY).

```

FT	ACT SITE	987	987	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	LIPID	2	2	MYRISTATE (POTENTIAL).
FT	DISULFID	199	212	BY SIMILARITY.
FT	DISULFID	206	225	BY SIMILARITY.
FT	DISULFID	219	236	BY SIMILARITY.
FT	DISULFID	659	671	BY SIMILARITY.
FT	DISULFID	666	684	BY SIMILARITY.
FT	DISULFID	678	693	BY SIMILARITY.
FT	DISULFID	788	912	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	826	842	BY SIMILARITY.
FT	DISULFID	926	993	BY SIMILARITY.
FT	DISULFID	957	972	BY SIMILARITY.
FT	DISULFID	983	1011	BY SIMILARITY.
FT	CARBOHYD	116	116	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	147	147	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	170	170	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	194	194	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	233	233	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	263	263	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	264	264	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	404	404	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	456	456	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	486	486	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	519	519	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	550	550	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	646	646	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	698	698	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	722	722	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	741	741	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	762	762	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	864	864	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	903	903	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	965	965	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPIC	166	192	Missing (in isoform Short). /FTid=VSP 005386.
FT	CONFLICT	808	808	R -> Y (IN REF. 3).
FT	SEQUENCE	1035	114887	AA: 114887 MW: E207970B08296E13 CRC64:

	Query Match	24.4%;	Score 571;	DB 1;	Length 1035;
	Best Local Similarity	35.0%;	Pred. No. 1.8e-41;		
	Matches 133;	Conservative	67;	Mismatches 138;	Indels 42;
				Gaps	15;
QY	69	IPRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDN	128		
Db	672	IPLVNLCDGFPHCKDGSDEAHCVRRLF-NGITDSSGLVQFR--IQSI-----WHVACAEN	722		
QY	129	FTEALAEATACRMGY-----SSKPTFRAVEIGDPDQDLDVVE-----ITENSQELMRNRSSGP	180		
Db	723	WTTQISDDVCQLLGLGTGNSSVPTF-STGGGPYVNLNTAPNGSLILTPSQQ-----	772		
QY	181	CLSGSLVSLHC--LACGKSLKT-----PRVVGGEASVDSWPWQVSIQYDKQHVCGGSILD	234		
Db	773	CLEDSLILLQCNKYKCGKLVTOEVSPIKIVGSDSREGAWPWVALYFDDQVCGASLVS	832		
QY	235	PHWLVTAHC-FRKHTDVFNWKVRAGSKLGSF--PSLAVAKIIIEFNPMY---PKDND	288		
Db	833	RDWLVSAAHCVYGRNMEPSKWKAVLGLHMASNLTSPOIETRLIDQIVINPHYNKRKQND	892		
QY	289	IALMKLQPLTFSGTVRPICLPFFDEELTPATPLWIIGWGFTKQGGKMSDILLQASVQV	348		
Db	893	IAMHLEMKVNYTDYIQICLPEENVQVPPGRICSIAGWGALIYQ-GSTADVLQEAADVPL	951		
QY	349	IDSTRCNADDAYQGEVTEKQMCAGIPEGGVDTCCQDGGGLMYQ-SDQWHVVGVISWGYG	407		
Db	952	LSNEKCY-QQMPENYNITENMVCAGYEAGVDSQGDSCGGLMCOENRNLLAGVTSFGYQ	1010		
QY	408	CGGPSTPGVYTKVSAYLNWI	427		
Db	1011	CALPNRPGVIARVPRFTEWI	1030		

RESULT 12
HEPS RAT

InterPro; IPR002172; LDL_receptor_A.
InterPro; IPR000998; MAM_domain.
InterPro; IPR000082; SEA_domain.
InterPro; IPR001254; Ser_protease_Try.
InterPro; IPR001190; Sscr_receptor.
Pfam; PF00431; CUB; 2.
Pfam; PF00057; ldl_recept_a; 2.
Pfam; PF00629; MAM; 1.
Pfam; PF01390; SEA; 1.
Pfam; PF00530; SRCR; 1.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
PRINTS; PR00261; LDLRECEPTOR.
PRINTS; PR00020; MAMDOMAIN.
SMART; SM00042; CUB; 2.
SMART; SM00192; LDLa; 2.
SMART; SM00137; MAM; 1.
SMART; SM00200; SEA; 1.
SMART; SM00202; SR; 1.
SMART; SM00020; Tryp_Spc; 1.
PROSITE; PS01180; CUB; 2.
PROSITE; PS01209; LDLRA_1; 2.
PROSITE; PS00688; LDLRA_2; 2.
PROSITE; PS00740; MAM_1; 1.
PROSITE; PS00600; MAM_2; 1.
PROSITE; PS00024; SEA; 1.
PROSITE; PS00420; SRCR_1; FALSE_NEG.
PROSITE; PS00287; SRCR_2; 1.
PROSITE; PS00240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SER; 1.
Signal-anchor; Glycoprotein; Hydrolase; Serine protease; Zymogen;
Transmembrane; Repeat.
CHAIN 1 829 NON-CATALYTIC CHAIN (HEAVY CHAIN).
CHAIN 830 1069 CATALYTIC CHAIN (LIGHT CHAIN).
DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
DOMAIN 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
TRANSMEM (POTENTIAL).
DOMAIN 48 1069 EXTRACELLULAR (POTENTIAL).
DOMAIN 52 169 SEA.
DOMAIN 227 268 LDL-RECEPTOR CLASS A 1.
DOMAIN 270 379 CUB 1.
DOMAIN 387 549 MAM.
DOMAIN 569 679 CUB 2.
DOMAIN 686 724 LDL-RECEPTOR CLASS A 2.
DOMAIN 723 816 SRCR.
DOMAIN 830 1069 SERINE PROTEASE.
ACT_SITE 874 874 CHARGE RELAY SYSTEM (BY SIMILARITY).
ACT_SITE 925 925 CHARGE RELAY SYSTEM (BY SIMILARITY).
ACT_SITE 1021 1021 CHARGE RELAY SYSTEM (BY SIMILARITY).
DISULFID 229 242 BY SIMILARITY.
DISULFID 236 255 BY SIMILARITY.
DISULFID 249 266 BY SIMILARITY.
DISULFID 688 700 BY SIMILARITY.
DISULFID 695 713 BY SIMILARITY.
DISULFID 707 722 BY SIMILARITY.
DISULFID 817 945 INTERCHAIN (BY SIMILARITY).
DISULFID 859 875 BY SIMILARITY.
DISULFID 959 1027 BY SIMILARITY.
DISULFID 991 1006 BY SIMILARITY.
DISULFID 1017 1045 BY SIMILARITY.
CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 380 380 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 579 579 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 675 675 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 727 727 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 751 751 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 770 770 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 791 791 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 936 936 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 999 999 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1069 AA; 118735 MW; E62549E463743C3D CRC64;
Query Match 23.6%; Score 552.5; DB 1; Length 1069;
Best Local Similarity 32.5%; Pred. No. 7.5e-40;
Matches 124; Conservative 71; Mismatches 150; Indels 37; Gaps 12;
QY 69 IPRKQLCDGELDCPLGEDEHCVKSPFEGPAVAVRLSKDRSTLQVLD-SATGNWFSACFD 127
Db IPLGNLSDSYPHCRDGSDEASCVRF-----LNGTRSNGLVQFNHSHIACAE 750
QY 128 NTFEALAEATACRMGYSSKPTFAVEI---GPDQLDVVVEITENSQELRMENSSGPPCLSG 184
Db NWTQTISNEVCHLLGLGSANSSMPISSTGGP-----FVRVNOAPNGSLILTPSLQCSQD 805
QY 185 SLVSLHC--LACGKSLKT----PRVVGEEASVDSWPWQVSIQYDKQH---VCGGSILD 234
Db SLILLQCNHKSCGEKKVTQKVPKIVGSDAQAGAWPWWVALYHRDRSTRLLCGASLVS 865
QY 235 PHWVLTAAHC-FRKHTDVFNNKVRAGSKLGSPSLAVAKII--IEFNPMYP---KDND 288
Db SDWLVSAAHCVYRRNLDPRWTAVLGLHMQSNLTSPQVRRVVDQIVINPHYDRRRKVD 925
QY 289 IALMKIQPLTFSGTVRPICLPFFDEELTPATPLMIIGWFTKQNGGKMSDILLQASQV 348
Db IAMHLEFKVNYTDYIQPICLPEENQIFIPGRTCISAGWYDKINAGSTVDVLKEADVPL 985
QY 349 IDSTECNADDAVQGEVTEKMKCAGIPEGGVDTCCGDSGGPLMYQ-SDQWHVVGIVSWGYG 407
Db ISNEKCO-QQLPEYNITESNICAGYEEGGIDSCQDSGGPLMCQENNRWFLVGVTSFGVQ 1044
QY 408 CGGSTPGVYTKVSAYLNWIYN 429
Db CALPNHGPVYVRVSQFIEWIHS 1066
RESULT 15
ENTK_PIG
ID -ENTK_PIG STANDARD; PRT; 1034 AA.
AC P98074;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
GN PRSS7 OR ENTK.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Duodenal mucosa;
RX MEDLINE=94327548; PubMed=8051081;
RA Matsushima M., Ichinose M., Yahagi N., Kakei N., Tsukada S.,
RA Miki K., Kurokawa K., Tashiro K., Shiohara K., Shinomiya K.,
RA Umeyama H., Inoue H., Takahashi T., Takahashi K.,
RT "Structural characterization of porcine enteropeptidase."
RL J. Biol. Chem. 269:19976-19982(1994).
CC -!- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in
CC trypsinogen.
CC -!- SUBUNIT: HETEROTRIMER OF A CATALYTIC (LIGHT) CHAIN, A MULTIDOMAIN
CC (HEAVY) CHAIN, AND A MINI CHAIN.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.

```

-!- PTM: THE MINI CHAIN MAY BE CLEAVED BY ELASTASE.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
-!- SIMILARITY: Contains 2 LDL-receptor class A domains.
-!- SIMILARITY: Contains 2 CUB domains.
-!- SIMILARITY: Contains 1 SEA domain.
-!- SIMILARITY: Contains 1 SRCR domain.
-!- SIMILARITY: Contains 1 MAM domain.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
-----
EMBL; D30799; BAA06459.1; -.
HSSP; P00763; LDPO.
MEROPS; S01.156; -.
InterPro; IPR001314; Chymotrypsin.
InterPro; IPR000859; CUB domain.
InterPro; IPR002172; LDL_receptor_A.
InterPro; IPR000998; MAM_domain.
InterPro; IPR000082; SEA_domain.
InterPro; IPR001254; Ser_protease_Try.
InterPro; IPR001190; Srcr_receptor.
Pfam; PF00431; CUB; 2.
Pfam; PF00057; ldl_recept_a; 2.
Pfam; PF00629; MAM; 1.
Pfam; PF01390; SEA; 1.
Pfam; PF00530; SRCR; 1.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
PRINTS; PR00261; LDLRECEPTOR.
PRINTS; PR00020; MAMDOMAIN.
SMART; SM00042; CUB; 2.
SMART; SM00192; LDLa; 2.
SMART; SM00137; MAM; 1.
SMART; SM00200; SEA; 1.
SMART; SM00202; SR; 1.
SMART; SM00020; Tryp_SPC; 1.
PROSITE; PS01180; CUB; 2.
PROSITE; PS01209; LDLRA_1; 2.
PROSITE; PS50068; LDLRA_2; 2.
PROSITE; PS00740; MAM_1; 1.
PROSITE; PS50060; MAM_2; 1.
PROSITE; PS50024; SEA; 1.
PROSITE; PS00420; SRCR_1; FALSE_NEG.
PROSITE; PS50287; SRCR_2; 1.
PROSITE; PS50240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SER; 1.
Signal-anchor; Glycoprotein; Myristate; Hydrolase;
Serine protease; Zymogen; Transmembrane; Repeat.
CHAIN 52 117 NON-CATALYTIC M CHAIN (MINI CHAIN).
CHAIN 118 799 NON-CATALYTIC H CHAIN (HEAVY CHAIN).
CHAIN 800 1034 CATALYTIC L CHAIN (LIGHT CHAIN).
DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
DOMAIN 48 1034 EXTRACELLULAR (POTENTIAL).
DOMAIN 52 169 SEA.
DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.
DOMAIN 240 349 CUB 1.
DOMAIN 357 519 MAM.
DOMAIN 539 649 CUB 2.
DOMAIN 656 694 LDL-RECEPTOR CLASS A 2.
DOMAIN 693 786 SRCR.
DOMAIN 800 1034 SERINE PROTEASE.
ACT_SITE 840 840 CHARGE RELAY SYSTEM (BY SIMILARITY).
ACT_SITE 891 891 CHARGE RELAY SYSTEM (BY SIMILARITY).
ACT_SITE 986 986 CHARGE RELAY SYSTEM (BY SIMILARITY).
LIPID 2 2 MYRISTATE (POTENTIAL).

```

FT	DISULFID	199	212	BY SIMILARITY.
FT	DISULFID	206	225	BY SIMILARITY.
FT	DISULFID	219	236	BY SIMILARITY.
FT	DISULFID	658	670	BY SIMILARITY.
FT	DISULFID	665	683	BY SIMILARITY.
FT	DISULFID	677	692	BY SIMILARITY.
FT	DISULFID	787	911	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	825	841	BY SIMILARITY.
FT	DISULFID	925	992	BY SIMILARITY.
FT	DISULFID	956	971	BY SIMILARITY.
FT	DISULFID	982	1010	BY SIMILARITY.
FT	CARBOHYD	116	116	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	147	147	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	170	170	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	194	194	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	283	283	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	343	343	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	350	350	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	403	403	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	455	455	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	485	485	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	518	518	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	549	549	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	645	645	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	697	697	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	701	701	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	721	721	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	740	740	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	761	761	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	804	804	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	863	863	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	902	902	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	964	964	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE	1034 AA;	114776 MW;	0388C64CF64CC368 CRC64;

Query Match 23.6%; Score 552; DB 1; Length 1034;

Best Local Similarity 34.2%; Pred. No. 7.9e-40;

Matches 127; Conservative 69; Mismatches 141; Indels 34; Gaps 14;

QY	74	LCDGELDCPLGDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDNFTEAL	133
Db	676	LCDGFSCHKDGSDEAHCVR-FLNGTANNSSLVQFR--IQSI-----WHTACAENWTTQT	726
QY	134	AETACRQMGY---SSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPGCLSGSLVSL	189
Db	727	SDDVCQLLGLGTGNSMPPFSSGG-GP-----FVKLNTAPNGSLILTASEQCFEDSLILL	780
QY	190	HC--LACGKSL---KTPRVVGEEASVDSPWQVSIQYDKQHVCGGSILDPHWLTAH	243
Db	781	QCNHKSCGKKQVAQEVSPKIVGNDSSREGAWPWWVALYNGQLLCSGLSVSRDLVLSAAH	840
QY	244	C-FRKHTDVFNWKVRAGSKLGSF--PSLAVAKIIIEFNPMY---PKNDIALMKLQFP	297
Db	841	CVYGRNLEPSKWAILGLHMTSNLTSPQIVTRLIDEIVINPHYNRRRKDSIAMHLEFK	900
QY	298	LTFSGTVRPICLPFFDEELTPATPLWIIIGWGFTKQNGKMSDILLQASVQVIDSTRCNAD	357
Db	901	VNYTDYIQICLPEENQVFPFGPRICSIAGWGVYQGGSP-ADILQEAADVPLLSNEKCO-Q	958
QY	358	DAYQGEVTEKMMKAGIPEGGVDTCCQDSCGGPLM-QSDQWHVVGVISWGYCGGSPSTPGV	416
Db	959	QMPYNTENNMKAGYEBGGIDSCQDSCGGPLMCLENRLLAGVTSFGYQCALPNRPGV	1018
QY	417	YTKVSAYLNWI	427
Db	1019	YARVPKFTTEWI	1029

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 07:45:10 ; Search time 65.2201 Seconds
(without alignments)
1721.140 Million cell updates/sec

Title: US-09-607-745-2
Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVKPLRP.....VYTKVSAYLNWIYNWKAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- SPTREMBL_23:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archheap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2167	92.5	405	4	Q96E86
2	1844	78.7	435	11	Q8VCA5
3	704.5	30.1	471	11	Q8CFE0
4	695.5	29.7	581	4	Q9BYE2
5	687	29.3	453	11	Q8KIT0
6	686.5	29.3	537	4	Q9BYE1
7	686.5	29.3	767	13	Q9DGR2
8	681	29.1	453	11	Q8VDE0
9	670.5	28.6	492	4	Q96T73
10	646	27.6	490	11	Q920K3
11	636.5	27.2	445	11	Q8CJ17
12	631.5	27.0	455	11	Q8CDR0
13	612	26.1	371	11	Q8CJ16
14	550.5	23.5	799	11	Q9DBI0
15	545	23.3	777	11	Q8CAN9
16	536	22.9	417	11	Q8VHJ4

17	531	22.7	802	4	Q8IUE2	Q8iue2 homo sapien
18	531	22.7	811	4	Q8IU80	Q8iu80 homo sapien
19	527	22.5	279	11	Q9QZ74	Q9qz74 rattus norv
20	523	22.3	417	11	Q8VDV1	Q8vdrv1 mus musculu
21	523	22.3	417	11	Q8VHK8	Q8vhk8 mus musculu
22	504.5	21.5	329	6	Q9GL10	Q9gl10 ovis aries
23	504.5	21.5	1524	13	Q91674	Q91674 xenopus lae
24	502	21.4	638	11	Q8R0P5	Q8rop5 mus musculu
25	501.5	21.4	439	11	Q8BHM9	Q8bhm9 mus musculu
26	499.5	21.3	327	4	Q8NI71	Q8ni71 homo sapien
27	499.5	21.3	572	11	Q8BIK6	Q8bik6 mus musculu
28	498.5	21.3	377	6	P79343	P79343 bos taurus
29	498.5	21.3	417	11	Q8BZ10	Q8bz10 mus musculu
30	496	21.2	643	6	Q97506	Q97506 sus scrofa
31	493.5	21.1	421	11	Q60491	Q60491 cavia porce
32	488	20.8	423	11	Q8BM10	Q8bm10 mus musculu
33	488	20.8	812	11	Q9R0W3	Q9row3 rattus norv
34	486.5	20.8	416	11	Q8BZ30	Q8bz30 mus musculu
35	486.5	20.8	416	11	Q8BZ13	Q8bz13 mus musculu
36	486	20.8	389	13	Q9PVX7	Q9pvx7 xenopus lae
37	482.5	20.6	284	4	Q8NF86	Q8nf86 homo sapien
38	481	20.5	321	4	Q96RZ8	Q96rz8 homo sapien
39	479.5	20.5	845	13	Q9DGR1	Q9dgr1 xenopus lae
40	478.5	20.4	415	6	Q29015	Q29015 sus sp. pre
41	477	20.4	1379	5	Q9V4N6	Q9v4n6 drosophila
42	476	20.3	581	5	Q9XZM7	Q9xzm7 strongyloce
43	474	20.2	855	11	Q9JJI7	Q9jjj17 rattus norv
44	473.5	20.2	260	13	Q9W7P9	Q9w7p9 paralichthy
45	473.5	20.2	310	11	Q9QYZ9	Q9qyz9 mus musculu

ALIGNMENTS

RESULT 1
Q96E86
ID Q96E86 PRELIMINARY; PRT; 405 AA.
AC Q96E86;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to transmembrane protease, serine 4 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; BC012752; AAH12752.1; -.
DR HSSP; P00761; IAN1.
DR MEROPS; S01.034; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; Tryp_SPC; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane.
FT NON TER 1
SQ SEQUENCE 405 AA; 44474 MW; 951ACD52D9D48E04 CRC64;

Query Match 92.5%; Score 2167; DB 4; Length 405;
Best Local Similarity 100.0%; Pred. No. 1.9e-202;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 IIALSLASIIIVVLIKVLIDKYVFLCGQPLHFIPRKQICDGLDCPLGEDEHCVKS 93
DB 4 IIALSLASIIIVVLIKVLIDKYVFLCGQPLHFIPRKQICDGLDCPLGEDEHCVKS 63

QY 94 FPEGPAVAVLSKDRSTLQVLDLSATGNWFSACFDNFTEALAEATACROMGYSSKPTFRAVE 153
DB 64 FPEGPAVAVLSKDRSTLQVLDLSATGNWFSACFDNFTEALAEATACROMGYSSKPTFRAVE 123

QY 154 IGPDDLDVVEITENSQELMRNNSGCLSGSLVSLHCLACGKSLKTPRVVGGEEASVDS 213
DB 124 IGPDDLDVVEITENSQELMRNNSGCLSGSLVSLHCLACGKSLKTPRVVGGEEASVDS 183

QY 214 WPMQVSIQDKQHVCGGSILDPHWVLTAAHCFRKHTDVFNNKVRAGSDKLGSPFSLAVAK 273
DB 184 WPMQVSIQDKQHVCGGSILDPHWVLTAAHCFRKHTDVFNNKVRAGSDKLGSPFSLAVAK 243

QY 274 IIIIEFNPMYKDNIALMKLQPLTFSGTVRPICLPFFDELTPTPLWIIGWFTKQN 333
DB 244 IIIIEFNPMYKDNIALMKLQPLTFSGTVRPICLPFFDELTPTPLWIIGWFTKQN 303

QY 334 GGRMSDILLQASVQVIDSTRCNADDAVQGEVTEKMMCAGIPEGGVDTTCQDSDGGLMYQS 393
DB 304 GGRMSDILLQASVQVIDSTRCNADDAVQGEVTEKMMCAGIPEGGVDTTCQDSDGGLMYQS 363

QY 394 DQHVHVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 435
DB 364 DQHVHVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 405

RESULT 2
Q8VCA5 PRELIMINARY; PRT; 435 AA.

ID Q8VCA5
AC Q8VCA5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to transmembrane protease, serine 4 (Channel-activating protease 2).
EN TMPSR4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=22144321; PubMed=12149280;
RA Vuagniaux G., Vallet V., Jaeger N.F., Hummler E., Rossier B.C.;
RT "Synergistic Activation of ENaC by Three Membrane-bound Channel-activating Serine Proteases (mCAP1, mCAP2, and mCAP3) and Serum- and Glucocorticoid-regulated Kinase (Sgk1) in Xenopus Oocytes.";
RT J. Gen. Physiol. 120:191-201(2002).
RL -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC EMBL; BC021368; AAH21368.1; --
DR EMBL; AY043240; AAK85307.1; --
DR HSSP; P00761; 1AN1.
DR DR
DR MEPOPS; S01.034; --
DR MGD; MGI:2384877; Tmpress4.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.

DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane.
SQ SEQUENCE 435 AA; 47495 MW; DC52E45A43E01369 CRC64;

Query Match 78.7%; Score 1844; DB 11; Length 435;
Best Local Similarity 77.2%; Pred. No. 6.4e-171;
Matches 336; Conservative 43; Mismatches 56; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVKLRKPRIPMETFRKVGIPILIALSLASIIIVVLIKVLIDKYF 60
DB 1 MESDSGQPLNRRDIVFRKPRRPQETFKVGIPILIALSLALVIVALLIKVLIDKYF 60

QY 61 LCGQPLHFIPRKQICDGLDCPLGEDEHCVKSPFEGPAVAVLSKDRSTLQVLDLSATGN 120
DB 61 ICGSPLTFIQRGQLCDGHLDCASGEDEHCVKDFEKPQGVAVLSKDRSTLQVLDAAATGT 120

QY 121 WFSACFDNFTEALAEATACROMGYSSKPTFRAVEICPDODLDVVEITENSQELMRNNSGP 180
DB 121 WASVCFDNFTEALAEATACROMGYDSQPAFRAVEIRPDQNLPAQVVTGNSQELQVQNGSRS 180

QY 181 CLSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPMQVSIQDKQHVCGGSILDPHWVLT 240
DB 181 CLSGSLVSLRCLDCGKSLKTPRVVGGVEAPVDSWPMQVSIQYNKQHVCGGSILDPHWILT 240

QY 241 AAHCFRKHTDVFNNKVRAGSDKLGSPFSLAVAKIIIEFNPMYKDNIALMKLQPLTF 300
DB 241 AAHCFRKLYDSSWKVRAGSNILGNSPLPVAKIFIAEPNPLYPKEDIALVKLQMLTF 300

QY 301 SGTVRPICLPFFDELTPTPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAV 360
DB 301 SGSVERPICLPFSDEVLPATPVVWVIGWFTENGKMSDMLLQASVQVIDSTRCNAEDAY 360

QY 361 QGEVTEKMMCAGIPEGGVDTTCQDSDGGLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKV 420
DB 361 EGEVTAEMLCAGTPQGGKDTCCQDSDGGLMYHSDKMQVVGIVSWGHGCGGPGSTPGVYTKV 420

QY 421 SAYLNWIYNVWKAE 435
DB 421 TAYLNWIYNVRKSEM 435

RESULT 3
Q8CFE0 PRELIMINARY; PRT; 471 AA.

ID Q8CFE0
AC Q8CFE0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to mosaic serine protease (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC042878; AAH42878.1; --
KW Protease.
FT NON TER 1
SQ SEQUENCE 471 AA; 52535 MW; ED58CFE6B7C3BCC4 CRC64;

Query Match 30.1%; Score 704.5; DB 11; Length 471;
Best Local Similarity 36.2%; Pred. No. 8.3e-60;
Matches 153; Conservative 70; Mismatches 167; Indels 33; Gaps 9;

QY 25 ETRFKVGIPILIALSLASIIIVVLIKVLIDKYFLCG-----QPLHFIPIRKQL-CD 76
DB 65 ETQRQ--LPLIGCVILLTSLVISLILL-----FYFWRGHTGIKYKEPLESCPIHAVRCD 116

QY 77 GELDCPLGEDEEHCVKSPFPGPAVAVRLSKDRSTLQVLD SATGNWFSACFDNFTALAE 136
Db 117 GVVDCKMKSDELGC-----VRFDWDKSLUKVYSGSSGEWLPVCSSSNDTDSKR 165
QY 137 ACROMGYSSKPTTFRAVEIGPDQDLDDVVEITENSQELMRNSSGPGCLSGSLVSLHCLACGK 196
Db 166 TCQQLGFDS--AYRTEVAHRDITSSFLSEYNTTIQESLYRSQCPSRRYVSLQCSHCGL 223
QY 197 SLKTPRVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLTAAHCF--RKHTDVFNW 254
Db 224 RAMTGRIVGGALTSESKWPWQVSLHFGTTHICGGLTIDAQWVLTAAHCFVVTREKLLGW 283
QY 255 KVRAGSDKLGSPFSLAVAKIIIEFNPMYPKDN-DIALMKLQFPLTPSGTVRPICLPFFD 313
Db 284 KVIAGTSLNHLQPEAASISQIIINGNYTDEQDDYDIALIRLSKPLTSAHIHPACLPMHG 343
QY 314 EELTPATPLWIIGWGFTKQNGKMSDILLQASVOVIDSTRCNADDAVQGEVTEKMMCAGI 373
Db 344 QTFLNETCWTGFGKTKETDEKTSPLFLREVQVNLIDFKKNDYLVDSYLTPRMCMAGD 403
QY 374 PEGGVDTCQGDSCGGLM-YQSDQHVHVGIVSWGCGGPGSTPGVTVKVSAYLNWIYNWK 432
Db 404 LRGRDSCQGDSCGGLVCEQNNRWYLAGVTSWGTGCGQKPKPGVTVKVTVEVLPWIYRKME 463
QY 433 AEL 435
Db 464 SEV 466

RESULT 4.

Q9BYE2 ID Q9BYE2 PRELIMINARY; PRT; 581 AA.
AC Q9BYE2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Membrane-type mosaic serine protease.
GN MSPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=21167393; PubMed=11267681;
RA Kim D.R., Sharmin S., Inoue M., Kido H.;
RT "Cloning and expression of novel mosaic serine proteases with and
RT without a transmembrane domain from human lung."
RL Biochim. Biophys. Acta 1518:204-209(2001).
DR EMBL; AB048796; BAB39741.1; --
DR HSSP; P00763; 1DPO.
DR MEROPS; S01.087; --
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 581 AA; 62689 MW; 4DAEB24D7D5BA4A CRC64;

Query Match 29.7%; Score 695.5; DB 4; Length 581;
Best Local Similarity 35.4%; Pred. No. 8.4e-59;
Matches 152; Conservative 69; Mismatches 161; Indels 47; Gaps 11;

QY 22 IPMETFR--KVGIPILIALSLASIIIVVVLKVLDKYVFLCGQ-----PLH 67
Db 149 LPKFTWRGQKQLPLIGCVLLIALVLSLIL-----FQWQGTGIRHKEQRESCPKH 202
QY 68 FIPRKQLCDGELDCPLGEDEEHCVKSPFPGPAVAVRLSKDRSTLQVLD SATGNWFSACFD 127
Db 203 AV-----RCDGVVDCKLKSDELGC-----VRFDWDKSLUKIYSGSSHQWLPICSS 247
QY 128 NPTALAEATACROMGYSSKPTTFRAVEIGPDQDLDDVVEITENSQELMRNSSGPGCLSGSLV 187
Db 248 NWNDYSEKTCRQLGFES--AHRTEVAHRDFANSFSLRYNSTIQESLHRSHCPSQRYI 305
QY 188 SLHCLACGKSLKTPRVVGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLTAAHCF-- 245
Db 306 SLQCSHCGLRAMTGRIVGGALASDSKWPWQVSLHFGTTHICGGLTIDAQWVLTAAHCFPV 365
QY 246 RKHTDVFNWKVRAGSDKLGSPFSLA-VAKIIIEFNPMY---PKNDIALMKLQFPLTFS 301
Db 366 TREKVLGKWKVYAGTSLNHLQPEAASIAEII--NSNYTDEDDYDIALMRLSKPLTLS 422
QY 302 GTVRPICLPFFDEELTPATPLWIIGWGFTKQNGKMSDILLQASVOVIDSTRCNADDAVQ 361
Db 423 AHIHPACLPMEHQTFSLNETCWTGFGKTKRETDDKTSPLFLREVQVNLIDFKKNDYLVVD 482
QY 362 GEVTEKMMCAGIPEGGVDTCCQDSCGGLM-YQSDQHVHVGIVSWGCGGPGSTPGVTVKV 420
Db 483 SYLTPRMCMAGDLHGRDSCQGDSCGGLVCEQNNRWYLAGVTSWGTGCGQKPKPGVTVKV 542
QY 421 SAYLNWIYN 429
Db 543 TEVLPWIYS 551

RESULT 5

Q8K1T0 ID Q8K1T0 PRELIMINARY; PRT; 453 AA.
AC Q8K1T0;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Type II transmembrane serine protease.
GN TMPSRS3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=citb-cj7;
RA Guipponi M., Scamuffa N., Scott H.S., Antonarakis S.E.;
RT "Isolation of the mouse Tmprss3 genomic DNA sequence."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AJ429216; CAD22137.1; --
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS50068; LDLRA_2; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane.
SQ SEQUENCE 453 AA; 49491 MW; 1ABCBF10AF6E1EF6 CRC64;

DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS50068; LDLRA_2; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 492 AA; 53863 MW; 3ABA755BF276DADF CRC64;

Query Match 28.6%; Score 670.5; DB 4; Length 492;
Best Local Similarity 38.5%; Pred. No. 1.8e-56;
Matches 148; Conservative 59; Mismatches 128; Indels 49; Gaps 13;

QY 75 CDGELDCPLGEDEEHCHVKSPFEGPAVAVRLSKDRSTLQVLDSATGNWFSACFNFTAL 134
DDB 133 CDGVSHCPGGEDENRCVRLY--GP-----NFILQYSSQRKSWHPVCQDDWNNYG 181
QY 135 ETACRQMGYSKPTFRAVEIGPDQDLQDLDWEITENSQELMRNMSG-----PCL 182
DDB 182 RAACKDMGY--KNFYSSQ-----GIVDDSGSTSFMKLNTSAGNVDIYKKLYHSDACS 232
QY 183 SGLVSLHCLACGKSL--KTPRVYGGEEASVDSWPWQVSIQYDKQHVCGGSIIDPHWVL 239
DDB 233 SKAVVSLRCTACGVNLSSRQSRIVGGESALPGAWPWQVSLHVQNVHVCVGGSIITPEWIV 292
QY 240 TAAHCFRKH-TDVFNWVKRAGSKLGF---PSLAVAKIIIEFNPMY---PKNDIAL 291
DDB 293 TAAHCFRKH-TDVFNWVKRAGSKLGF---PSLAVAKIIIEFNPMY---SHPNYDSKTKNDIAL 348
QY 292 MKLQFPLTFSGTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDS 351
DDB 349 MKLQKPLTFNDLVKPVCLPNPGMMLQPEQLCWISGWGATEEK-GKTSEVLNAKVLLIET 407
QY 352 TRCNADDAVQGEVTEKMKCAGIPEGVDTCQDSDGGPLMTQSDQ-WHVVGIVSWGYCGG 410
DDB 408 QRCNSRYVDNLITPAMICAGFLQGNVDSCQDSDGGPLVTSKNNIWLIGDTSWGGCAK 467
QY 411 PSTPGVYTKVSAYLNWIYVNWKA 434
DDB 468 AYRPGVYGNVMTVDWIYRQMRAD 491

RESULT 10
Q920K3 PRELIMINARY; PRT; 490 AA.
ID Q920K3
AC Q920K3
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE TMPS2.
GN TMPS2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Tsuzuki S.;
RT "TMPS2, Rat.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AB073550; BAB70683.1; -.
DR HSP; P00761; 1AN1.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL receptor A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Srcr receptor.
Pfam; PF00057; ldl_recept_a; 1.

DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50068; LDLRA_2; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 490 AA; 53518 MW; 2BC691551CAC409A CRC64;

Query Match 27.6%; Score 646; DB 11; Length 490;
Best Local Similarity 38.8%; Pred. No. 4.4e-54;
Matches 151; Conservative 54; Mismatches 148; Indels 36; Gaps 13;

QY 62 CGQPLHFIPRKLQCDGELDCPLGEDEEHCHVKSPFEGPAVAVRLSKDRSTLQVLDSATGNW 121
DDB 119 CGSGTCTISSSLWCDGVAQCNGKDNRC-----VRLYGTSTLQVYSSQRKAW 167
QY 122 FSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQ-----DLDDVEITENSQEL-RMRN 176
DDB 168 YPVQDDWNESYGRAACKDMGY--KNSFYSSQGIPTDQSGATSEFMKLNVSAGNVLDLYKKLY 225
QY 177 SSGCLSGSLVSLHCLACG-KSL-KTPRVYGGEEASVDSWPWQVSIQYDKQHVCGGSIID 234
DDB 226 HSDSCSSRMVSLRCIECGVRSVRRQSRIVGGSTASPGDWQVSLHVQGHVCGGSIIT 285
QY 235 PHWVLTAAHCFRKH-TDVFNWVKRAGSKD-----LGSFPLAVAKIIIEFNPMY---PK 285
DDB 286 PEWIVTARHCVEEPLSSPRYWTAFAGILKQSLMFYGS--RHQVEKVI---SHPNYDSKTK 340
QY 286 DNDIALMKLOFPLTFSGTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLQAS 345
DDB 341 NNDIALMKLOFPLAFNDVVKVCLPNPGMMLDLAQECWISGWGATYEK-GKTSVDLNAAM 399
QY 346 VQVIDSTRCNADDAVQGEVTEKMKCAGIPEGVDTCQDSDGGPLM-YQSDQWHVVGIVSW 404
DDB 400 VPLIEPSKCNKSIYNNLITPAMICAGFLQGSVDSCQDSDGGPLVTLKNEIWWLIGDTSW 459
QY 405 GYGCGPSTPGVYTKVSAYLNWIYVNWKA 433
DDB 460 GSGCAKAYRPGVYGNVMTVDWIYQMR 488

RESULT 11
Q8CJ17 PRELIMINARY; PRT; 445 AA.
ID Q8CJ17
AC Q8CJ17
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Adrenal mitochondrial protease long variant.
GN AMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEDH;
RA Omer S.; Bicknell A.B.; Lowry P.J.;
RT "Identification of a rat adrenal mitochondrial protease.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF537098; AAN06757.1; -.
KW Protease.
SQ SEQUENCE 445 AA; 48440 MW; BE3F56D8372ED988 CRC64;

Query Match 27.2%; Score 636.5; DB 11; Length 445;
Best Local Similarity 36.2%; Pred. No. 3.3e-53;
Matches 135; Conservative 62; Mismatches 141; Indels 35; Gaps 10;

QY 79 LDCPLGEDEEHCHVKSPFEGPAVAVRLSKDRSTLQVLDSATGNWFSACFNFTALAEATAC 138

Db 81 LSCPGVSSEKLLPSLPK--AVSFRINGEDLLLEVOVRARPDWLLVCHEGWNPALGMHIC 138

QY 139 ROMGYSSKPTFRAVEIGPDQDLVVVEITENSQELMRNS-----SGPCLSGSL 186

Db 139 QSLGYFRLTQHKAVNL---SDIKL-----NRSQFAQLSARPGSLVEEAWQPSTNCPSGRI 191

QY 187 VSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLTAAHC-- 244

Db 192 VSLKCECGARPLASRIVGGQAVASGRWPWQASVMLGSRHTCGSVLAPYVWVTTAAHCY 251

QY 245 -FRKHTDVFNWVKVRAG---SDKLGSPFSLAVAKIIIEFNMPYKDN---DIALMKLQFP 297

Db 252 SFRL-SRLSSWRVHAGLVSHSAVRHQGTWVEKIIP---HPLYSAQNHDYDVALLQLRTP 307

QY 298 LTFSGTVRPICLPFFDEELTPATPLWIIGWFTKQNGGKMSDILLQASVQVIDSTRCNAD 357

Db 308 INFSDTVSAVCLPAKEOHFPQGSQCWVSGWGHTDPSHTSSDTLQDTMVPLLLSTDLNCS 367

QY 358 DAYQGEVTEKMMKAGIPEGGVDTCCQDSGGPLMYQS-DQWVVVGIVSWGVCSTPSTPGV 416

Db 368 CMYSGALTHRMCLCAGYLDGRADACQDSGGPLVCPSGDTWHLVGVVSWGRGCAEPNRP 427

QY 417 YTKVSAYLNWIYN 429

Db 428 YAKVAEFLDWIHD 440

RESULT 12

Q8CDR0

D Q8CDR0 PRELIMINARY; PRT; 455 AA.

AC Q8CDR0;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

TT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

T 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

ES Transmembrane protease.

SS Mus musculus (Mouse).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

C NCBI_TaxID=10090;

X [1]

N SEQUENCE FROM N.A.

P STRAIN=C57BL/6J; TISSUE=Testis;

C MEDLINE=22354683; PubMed=12466851;

X The FANTOM Consortium,

A the RIKEN Genome Exploration Research Group Phase I & II Team;

A "Analysis of the mouse transcriptome based on functional annotation of

T 60,770 full-length cDNAs."

T Nature 420:563-573 (2002).

R EMBL; AK029714; BAC26577.1; ..

Q SEQUENCE 455 AA; 49669 MW; BE22EB2E7503C74B CRC64;

Query Match 27.0%; Score 631.5; DB 11; Length 455;

Best Local Similarity 36.2%; Pred. No. 1e-52;

Matches 135; Conservative 63; Mismatches 140; Indels 35; Gaps 10;

Y 79 LDCPLGEDEEHCVKSPFEGPAPAVARLSKDRSTLQVLDSATGNWFSACFDNFTALAEATAC 138

b 91 LNCPGVSREEELLPSLPK--TVSFRINGEDLLLVQVRARPDWLLVCHEGWSPALGMHIC 148

Y 139 ROMGYSSKPTFRAVEIGPDQDLVVVEITENSQELMRNS-----SGPCLSGSL 186

b 149 KSLGHIRLTQHKAVNL---SDIKL-----NRSQFAQLSARPGGLVEESWKPSANCPSGRI 201

Y 187 VSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLTAAHC-- 244

b 202 VSLKCECGARPLASRIVGGQAVASGRWPWQASVMLGSRHTCGSVLAPHWVTTAAHCY 261

Y 245 -FRKHTDVFNWVKRAGSKLGS---FPSLAVAKIIIEFNMPYKDN---DIALMKLQFP 297

b 262 SFRL-SRLSSWRVHAGLVSHGAVRQHQGTWVEKIIP---HPLYSAQNHDYDVALLQLRTP 317

Y 298 LTFSGTVRPICLPFFDEELTPATPLWIIGWFTKQNGGKMSDILLQASVQVIDSTRCNAD 357

Db 318 INFSDTVGAVCLPAKEOHFPWGSQCWVSGWGHTDPSHTSSDTLQDTMVPLLLSTLCNSS 377

QY 358 DAYQGEVTEKMMKAGIPEGGVDTCCQDSGGPLMYQS-DQWVVVGIVSWGVCSTPSTPGV 416

Db 378 CMYSGALTHRMCLCAGYLDGRADACQDSGGPLVCPSGDTWHLVGVVSWGRGCAEPNRP 437

QY 417 YTKVSAYLNWIYN 429

Db 438 YAKVAEFLDWIHD 450

RESULT 13

Q8CU16

ID Q8CU16 PRELIMINARY; PRT; 371 AA.

AC Q8CU16;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Adrenal mitochondrial protease short variant.

GN AMP.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NEDH;

RA Omer S.; Bicknell A.B.; Lowry P.J.;

RT "Identification of a rat adrenal mitochondrial protease."

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF537099; AAN06758.1; ..

KW Protease.

SQ SEQUENCE 371 AA; 40694 MW; 89A64081D9A1FE26 CRC64;

Query Match 26.1%; Score 612; DB 11; Length 371;

Best Local Similarity 34.7%; Pred. No. 6.2e-51;

Matches 131; Conservative 62; Mismatches 138; Indels 46; Gaps 11;

QY 87 BEHCVKSPFEGPA-----VAVRLSKDRSTLQVLDSATGNWFSACFDNFTAL 134

Db 2 EIRCTEEGP-GPGIFRAELGDDQQQPIISFRINGEDLLLEVOVRARPDWLLVCHEGWNPALG 60

QY 135 ETACRQMGYSSKPTFRAVEIGPDQDLVVVEITENSQELMRNS-----SGPCL 182

Db 61 MHICQSLGYFRLTQHKAVNL---SDIKL-----NRSQFAQLSARPGSLVEEAWQPSTNCP 113

QY 183 SGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWVLTAA 242

Db 114 SGRIVSLKCECGARPLASRIVGGQAVASGRWPWQASVMLGSRHTCGSVLAPYVWVTTAA 173

QY 243 HC---FRKHTDVFNWVKVRAG---SDKLGSPFSLAVAKIIIEFNMPYKDN---DIALMK 293

Db 174 HCMYSFRL-SRLSSWRVHAGLVSHSAVRHQGTWVEKIIP---HPLYSAQNHDYDVALLQ 229

QY 294 LQFPLTFSGTVRPICLPFFDEELTPATPLWIIGWFTKQNGGKMSDILLQASVQVIDSTR 353

Db 230 LRTPIINFSDTVSAVCLPAKEOHFPQGSQCWVSGWGHTDPSHTSSDTLQDTMVPLLLSTDL 289

QY 354 CNADDAYQGEVTEKMMKAGIPEGGVDTCCQDSGGPLMYQS-DQWVVVGIVSWGVCSTPSTPGV 412

Db 290 CNSSCMYSGALTHRMCLCAGYLDGRADACQDSGGPLVCPSGDTWHLVGVVSWGRGCAEPN 349

QY 413 TPGVYTKVSAYLNWIYN 429

Db 350 RPYAKVAEFLDWIHD 366

RESULT 14

Q9DBI0

ID Q9DBI0 PRELIMINARY; PRT; 799 AA.

AC Q9DBI0;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

```

Db      541  --GQSDCRDGS-DEQHC-DCGLQGLSSRIVGGTVSSGEWPQASLQIRGRHICGGALLA 596
Qy      235  PHWVLTAAHCFERKHT-----DVFNWVKVRAGSKLGSFPSPSLAVAKIIIEFNPMYPKDN 287
Db      597  DRWVITAACHCFQEDSMASPKLWTVFLGKVRQNSRWPGEV-SFKVSRFLFHPYHEEDSHDY 655
Qy      288  DIALMKLQFPLTFSGTVRPICLPFFDEELTPATPLWIIIGWGTQKQNGKMSDILLQASVQ 347
Db      656  DVALLQLDHPVVSATVRPVCLPARSHFFEPGQHCWITGWG-AQREGGPPVNTLOKVDVQ 714
Qy      348  VIDSTRCNADDAQGEVTEKMMKAGIPEGGVDTCCGDSGGPLMYQ--SDQHVVGIVSWG 405
Db      715  LVPQDLCS--EAYRQVSPRMLCAGYRKGKKDACQGDSCGGPLVCREPSGRWFLAGLVSQ 772
Qy      406  YGCGGPGSTPGVYTKVSAYLNIYNV 430
Db      773  LGCGRPNFFGVYTRVTRVINWIOQV 797

RESULT 15
Q8CAN9
ID Q8CAN9 PRELIMINARY; PRT; 777 AA.
AC Q8CAN9;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Protease (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK038356; BAC29973.1; -.
FT NON TER 1
SQ SEQUENCE 777 AA; 87314 MW; A18E2F4ECF06D3A8 CRC64;

Query Match 23.3%; Score 545; DB 11; Length 777;
Best Local Similarity 32.5%; Pred. No. 5,9e-44;
Matches 124; Conservative 72; Mismatches 148; Indels 38; Gaps 13

Qy      69  IPRKQLCDGELDCPLGDEEHCVKSFPEGPAVAVRLSKDRSTLOVLD-SATGNWFSACFD 127
Db      410  IPLGNLCDSPYHCRDGSDEASCVRF-----LNGTRSNNGLVFNHSHIWHIACAE 459
Qy      128  NFEALAEACROMGYSSKPTTFAVEI---GPDQDLDVVEITENSQELRMNSSGPCLSG 184
Db      460  NWTQISNEVCHLLGLGSANSSMPISTGGP-----FVRVNAQPNGLILTPSLQCSQD 514
Qy      185  SLVSLHC--LACGKSLKT----PRVVGEEASVDSPWQVSIQYDKQH---VCGGSILD 234
Db      515  SLILLQCNHKSCGEKKVTQKVPKIVGGSDAQAGAPWVVALYHRDRSTDRLLCGASLVS 574
Qy      235  PHWVLTAAHC-FRKHTDVFNWVKVRAGSKLGSFPSPSLAVAKII--IEFNPMYP---KND 288
Db      575  SDWLVSAAHCVYRRNLDPTRWTAVLGLHMQSNLTSPQVVRVVDQIVNPHYDRRRKVD 634
Qy      289  IALMKLQFPLTFSGTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQ 348
Db      635  IAMHLEFKVNYTDYIQIPLPEENQIFIPGRTCSIAGWGYDKINGSTV-DVLKEADVPL 693
Qy      349  IDSTRCNADDAQGEVTEKMMKAGIPEGGVDTCCGDSGGPLMYQ--SDQHVVGIVSWG 407
Db      694  ISNEKCO-QQLPEYNITESMICAGYEKGIDSCQDSCGGPLCMCQENNRNFWLGVTSFGVQ 752

```

QY 408 CCGPSTPGVYTKVSAYLNWIYN 429
| | : | | | : | | : | | :
Db 753 CALPNHPGVYVRVSOPIEWIHS 774

Search completed: November 23, 2003, 07:51:54
Job time : 71.2201 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

MM protein - nucleic search, using frame_plus_p2n model

Run on: November 23, 2003, 07:52:05 ; Search time 5493.45 Seconds
(without alignments)
3239.440 Million cell updates/sec

Title: US-09-607-745-2
Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVKPLRKP.....VYTKVSAYLNWIYVWKAEL 435

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
MODEL=frame+p2n.model -DEV=xlh
Q=/cgn2_1/USPTO.spool/US09607745/runat 21112003 144345 22184/app query.fasta_1.1038
DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
USER=US09607745.@CGN_1_1_5308@runat 21112003 144345 22184 -NCPU=6 -ICPU=3
NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_nam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2342	100.0	1305	6	AX076190	AX076190 Sequence
2	2338	99.8	2038	6	AR142620	AR142620 Sequence
3	2338	99.8	2038	6	BD137129	BD137129 Human pro
4	2337	99.8	2081	9	AF179224	AF179224 Homo sapi
5	2337	99.8	2088	9	BC011703	BC011703 Homo sapi
6	2329	99.4	1479	6	AX076192	AX076192 Sequence
7	2324	99.2	2079	6	AR232520	AR232520 Sequence
8	2324	99.2	2079	6	AX207967	AX207967 Sequence
9	2324	99.2	2079	9	AF216312	AF216312 Homo sapi
10	2319	99.0	2137	6	AX207899	AX207899 Sequence
11	2297.5	98.1	2063	6	AX092380	AX092380 Sequence
12	2297.5	98.1	2063	6	AX376262	AX376262 Sequence
13	2297.5	98.1	2063	6	AX395214	AX395214 Sequence
14	2297.5	98.1	2063	6	AX697206	AX697206 Sequence
15	2167	92.5	1793	9	BC012752	BC012752 Homo sapi
16	1844	78.7	2060	10	BC021368	BC021368 Mus muscu
17	1844	78.7	2268	10	AY043240	AY043240 Mus muscu
18	1287	55.0	699	6	E13203	E13203 Human cDNA
19	704.5	30.1	2992	10	BC042878	BC042878 Mus muscu
20	699.5	29.9	2067	6	AX149581	AX149581 Sequence
21	697.5	29.8	1314	6	AX190543	AX190543 Sequence
22	697.5	29.8	1314	6	AX190546	AX190546 Sequence
23	697.5	29.8	1689	6	AX149579	AX149579 Sequence
24	697.5	29.8	1748	6	AX369323	AX369323 Sequence
25	694	29.6	609	6	AX076194	AX076194 Sequence
26	692.5	29.6	2393	9	AB048796	AB048796 Homo sapi
27	690.5	29.5	2237	9	AY190317	AY190317 Homo sapi
28	687.5	29.4	3222	9	BC018715	BC018715 Homo sapi
29	686.5	29.3	3507	5	AB038497	AB038497 Xenopus l
30	683.5	29.2	1614	6	AX360089	AX360089 Sequence
31	683.5	29.2	3324	9	AB048797	AB048797 Homo sapi
32	683	29.2	2752	10	MMU300738	AJ300738 Mus muscu
33	677.5	28.9	2479	6	AR225372	AR225372 Sequence
34	677.5	28.9	2479	6	AX201121	AX201121 Sequence
35	677.5	28.9	2479	6	AX207965	AX207965 Sequence
36	677.5	28.9	2479	6	AX267920	AX267920 Sequence
37	677.5	28.9	2479	6	AX306771	AX306771 Sequence
38	677.5	28.9	2479	6	AX395318	AX395318 Sequence
39	677.5	28.9	2479	9	HSU75329	U75329 Human serin
40	677.5	28.9	3966	6	AX041973	AX041973 Sequence
41	676.5	28.9	1479	6	AR225345	AR225345 Sequence
42	676.5	28.9	1738	6	AX395316	AX395316 Sequence
43	676.5	28.9	1740	9	AF270487	AF270487 Homo sapi
44	676.5	28.9	3195	9	BC051839	BC051839 Homo sapi
45	675.5	28.8	1476	6	AX201158	AX201158 Sequence

ALIGNMENTS

```
AX076190
LOCUS AX076190 1305 bp DNA linear PAT 06-FEB-2001
DEFINITION Sequence 1 from Patent WO0104141.
ACCESSION AX076190
VERSION AX076190.1 GI:12710815
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Suendermann,B., Hofmann,U., Matzku,S. and Wilbert,O.
TITLE Seripancrin
JOURNAL Patent: WO 0104141-A 1 18-JAN-2001;
MERCK PATENT GmbH (DE)
FEATURES
     source
     1..1305
     /organism="Homo sapiens"
     /mol_type="genomic DNA"
     /db_xref="taxon:9606"
     1..>1305
     /note="unnamed protein product"
     /codon_start=1
     /protein_id="CAC28479.1"
     /db_xref="GI:12710816"
     /translation="MDPDSQPLNSLDVVKPLRPIMETFRKVGIPITALLSLASI
IIVVLKIVLDKYLFGQPLHFIPKQLCDGELDCPGEDEEHCVKSFEGPAVAV
RLSKDRSTLOVLSATGNWFSACFDNTEALAEATACROMGYSSKPTFRAVEIGPDQDL
DVEITSTNSQELMRNSSGPGCLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWFWQV
SIQYDKQHVCGSILDPHVLVTAACFRKHTDVFNKVRAGSKLGSFPLSLAVAKII
TEFNPMYKNDIALMKLQPLTFSGIVRPICLPFEDELTPTPLWIIIGWFTKQNG
GKMSDILLQASVQVIDSTRCNADDAVQGEVTEKMMKAGIPEGGVDTCCQSDSGPLMYQ
SDQWHVGVISWVGCGGPGSTPGVYTKVSAYLNWIYNVWKABL"
BASE COUNT 293 a 368 c 372 g 272 t
ORIGIN
Alignment Scores:
pred. No.: 9,52e-215 Length: 1305
Score: 2342.00 Matches: 435
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-607-745-2 (1-435) x AX076190 (1-1305)
Y 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
b 1 ATGGATCCCTGACAGTGATCAACCTCTGAACAGCCCTCGATGTCAAACCCCTGCGCAACCC 60
Y 21 ArgileProMetGluThrPheArgLysValGlyIleProIleIleLeuAlaLeuLeuSer 40
b 61 CGTATCCCATGGAGACCTTTCAGAAAGGTGGGGATCCCATCATCATAGCACTACTGAGC 120
Y 41 LeuAlaSerIleIleValValValLeuIleValLysValIleLeuAspLysTyrTyrPhe 60
b 121 CTGGCGAGTATCATATTGTGGTTCTCTCATCAAGGTGATTCCTCATCAAGTAAATACTACTTC 180
Y 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
b 181 CTCTCGGGGAGAGCTCTCCACTTCATCCCGAGGAGAGCTGTGTGACGGAGAGCTGGAC 240
Y 81 CysProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaVal 100
b 241 TGTCCCTTGGGGAGGAGGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCGCTGCAGTG 300
Y 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
b 301 GCAGTCCGCTCTCCAAAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGGAAC 360
Y 121 TrpPheSerAlaCysPheAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
b 361 TGGTCTCTGCCTGTTCGACAACTTCACAGAGCTCTCCTGCTGAGACAGCTGTAGGCAG 420
```

```
QY 141 MetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
Db ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGAGATTGGCCACAGCAGGATCTG 480
QY 161 AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180
Db GATGTTGTTGAATCACAGAAAACAGCCAGGAGCTTCGCATCGGAACCTCAAGTGGGCC 540
QY 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
Db TGTCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTCCTGTGGAGAGCCTGAAGACC 600
QY 201 ProArgValValGlyGlyGluGluAlaSerValAspSerTyrProTyrGlnValSerIle 220
Db CCCCCTGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTCAAGCATC 660
QY 221 GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTyrValLeuThr 240
Db CAGTACGACAAACAGCACGCTCTGTGGAGGAGCATCTCTGGACCCCTGAGCTCCTCACG 720
QY 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTyrLysValArgAlaGlySer 260
Db GCAGCCCACTGCTTTCAGGAAACATACCGATGTGTTCATGAAAGTGGGCGAGGCTCA 780
QY 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsn 280
Db GACAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATCAATTCAATCAAC 840
QY 281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
Db CCCATGTACCCCAAGACAAATGACATCGCCCTCATGAAGTGCAGTTCCTCCACTCACTTC 900
QY 301 SerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThr 320
Db TCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGTCACTCCAGCCACC 960
QY 321 ProLeuTyrIleIleGlyTyrGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle 340
Db CCCTCTGGATCATTTGGATGGGCTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATA 1020
QY 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyr 360
Db CTGCTGCAGGCGTCAAGTCCAGGTCAATGACAGCACACGGTGTCAATGCAGACGATCGGTAC 1080
QY 361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr 380
Db CAGGGGGAAGTCAACGAGAAGATGATGTGTGAGGATCCCGGAAGGGGTGTGACACACC 1140
QY 381 CysGlnGlyAspSerGlyProLeuMetTyrGlnSerAspGlnTyrHisValValGly 400
Db TGCAGGGTGACAGTGGTGGGCTATGCTGTGAGGATCCCGGAAGGGGTGTGAGGAGC 1200
QY 401 IleValSerTyrGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVal 420
Db ATCGTTAGTGGGCTATGCTGTGGGCGCCCTGATGTACCAATCTGACAGTGGCATGTGGTGGC 1260
QY 421 SerAlaTyrLeuAsnTyrIleTyrAsnValTyrLysAlaGluLeu 435
Db TCAGCTATCTCACTGAGTGTACATGTCTGGAAGGCTGAGCTG 1305
RESULT 2
AR142620
LOCUS AR142620 2038 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 18 from patent US 6203979.
ACCESSION AR142620
VERSION AR142620.1 GI:15103906
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2038)
AUTHORS Bandman,O., Hillman,J.L., Yue,H., Guegler,K.J., Corley,N.C.,
```


Tang, Y. Tom, and Shah, P.
Human protease molecules
Patent: US 6203979-A 18 20-MAR-2001;
Location/Qualifiers
1. .2038
/organism="unknown"

BASE COUNT 462 a 591 c 569 g 416 t
ORIGIN

Alignment Scores:
Pred. No.: 4.11e-214 Length: 2038
Score: 2338.00 Matches: 434
Percent Similarity: 100.00% Conservative: 1
Best local Similarity: 99.77% Mismatches: 0
Query Match: 99.83% Indels: 0
DB: 6 Gaps: 0

US-09-607-745-2 (1-435) x AR142620 (1-2038)

QY 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspVallysProLeuArgLysPro 20
DB 200 ATGGATCCTGACAGTGAACCTCTGAACAGCCTCGATGTCAAAACCCCTGGCGAAACCC 259
QY 21 ArgIleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSer 40
DB 260 CGTATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 319
QY 41 LeuAlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrPhe 60
DB 320 CTGGCGAGTATCATATTGTGGTGTCTCTCAAGGTGATTCCTGGATAAATACTACTTC 379
QY 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
DB 380 CTCCTGGGGCAGCCTCTCCACTTCATCCCGAAGCAGCTGTGTGACGGAGCTGGAC 439
Y 81 CysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaVal 100
DB 440 TGTCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTGCAGTG 499
Y 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
DB 500 GCAGTCCGCTCTCCAAAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGGAAC 559
Y 121 TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
DB 560 TGGTTCTCTGCCTGTTTCGACAACTTCACAGAACTCTCGCTGAGACAGCCTGTAGGCG 619
Y 141 MetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
DB 620 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCAGACAGGATCTG 679
Y 161 AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180
DB 680 GATGTTGTTGAATACAGAAACACAGAAACAGCCAGGAGCTTCGATCGGAACTCAAGTGGGCC 739
Y 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
DB 740 TGTCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTTGGCTGTGGGAGAGCCTGAAGACC 799
Y 201 ProArgValValGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
DB 800 CCCCGTGTGGTGGGAGGAGGCTCTGTGATCTTGGCTTGGCTGGCAGGTCAGCATC 859
Y 221 GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
DB 860 CAGTACGACAAACAGCACGCTCTGTGGAGGAGCATCTTGACCCCTGCTGGTCTCAGC 919
Y 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260
DB 920 GCAGCCCACTGCTTCAGGAACATACCGATGTGTCACTGGAAGGTGCGGGCAGGCTCA 979
Y 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsn 280

980 GACAAACTGGGCGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTAATCAAC 1039
QY 281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
DB 1040 CCCATGTATCCCAAGACATGACATGCCCTCATGAAGTGCAGTCCCACTCACTTTC 1099
QY 301 SerGlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThr 320
DB 1100 TCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 1159
QY 321 ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIle 340
DB 1160 CCACTCTGGATCATTTGATGGGCTTTACGAAGCAGAATGGAGGAAGATGTCTGACATA 1219
QY 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyr 360
DB 1220 CTGCTGCAGGCTCAGTCCAGTCAATGACAGCACAGGTGCAATGCAGACGATCGGTAC 1279
QY 361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr 380
DB 1280 CAGGGGGAAGTCACCGAGAAGATGATGTGTGACAGCATCCCGGAAGGGGTGTGGACACC 1339
QY 381 CysGlnGlyAspSerGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGly 400
DB 1340 TGCCAGGCTGACAGTGTGGGCCCTGATGTACCAATCTGACCACTGGCATGTGGTGGC 1399
QY 401 IleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVal 420
DB 1400 ATCGTTAGCTGGGCTATGGCTGGGGGGCCCGAGCACCCAGAGTATACACCAAGGTC 1459
QY 421 SerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
DB 1460 TCAGCCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1504

RESULT 3
BD137129
LOCUS Human protease molecule.
DEFINITION Human protease molecule.
ACCESSION BD137129
VERSION BD137129.1 GI:23232074
KEYWORDS JP 2002508970-A/6.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2038)
AUTHORS Bandman, O., Hillman, J.L., Yue, H., Guegler, K.J., Corley, N.C.,
Tang, T.Y. and Shah, P.
TITLE Human protease molecule
JOURNAL Patent: JP 2002508970-A 6 26-MAR-2002;
COMMENT INCYTE PHARMACEUTICALS INC
OS Homo sapiens (human)
PN JP 2002508970-A/6
PD 26-MAR-2002
PF 12-JAN-1999 JP 2000540252
PR 16-JAN-1998 US 09/008271
PI OLGA BANDMAN, JENNIFER L HILLMAN, HENRY YUE, KARL J GUEGLER, NEIL
PI C CORLEY.
PI TOM Y TANG, PURVI SHAH
PC C12N15/09, A61K38/46, C07K16/40, C12N1/19, C12N1/21, C12N5/10 PC
, C12N9/48, C12N9/64,
PC C12Q1/68, C12N15/00, A61K37/54, C12N5/00
CC 1337018
FH Key
FT source
FT Location/Qualifiers
1. .2038
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source
1. .2038
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 462 a 591 c 569 g 416 t
ORIGIN

Alignment Scores:
 Pred. No.: 4,11e-214 Length: 2038
 Score: 2338.00 Matches: 434
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.77% Mismatches: 0
 Query Match: 99.83% Indels: 0
 DB: 6 Gaps: 0

US-09-607-745-2 (1-435) x BD137129 (1-2038)

QY 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
 DB 200 ATGGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAACCC 259
 QY 21 ArgIleProMetGluThrPheArgLysValGlyIleProIleIleIleIleAlaLeuLeuSer 40
 DB 260 CGTATCCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTAGC 319
 QY 41 LeuAlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrPhe 60
 DB 320 CTGGCGAGTATCATCATTTGGTTTCTCTCATCAAGGTGATCTGGATAAATACTACTTC 379
 QY 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
 DB 380 CTCTCGGGCAGCCTCTCCACTTTCATCCCGAGGAGCAGCTGTGTGACGGAGCTGGAC 439
 QY 81 CysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaVal 100
 b 440 TGTCCCTTGGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTGCAGTG 499
 Y 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
 b 500 GCAGTCCGCTCTCCAAAGGACCGATCCACACTGCAGGTGTGACTCGGCCACAGGGAAC 559
 Y 121 TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
 b 560 TGGTTCTCTGCTCTTTCGACAACTTCACAGAAGCTCTCGGTGAGACAGCTGTAGGCAG 619
 Y 141 MetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
 b 620 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGAGATTGGCCAGACCCAGATCTG 679
 Y 161 AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180
 b 680 GATGTTGTTGAATACAGAAACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCC 739
 Y 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
 b 740 TGTCTCTCAGGCTCCCTGGTCTCCCTGCAGTGTCTTGCCTTGGGAGAGCCTGAAGACC 799
 Y 201 ProArgValValGlyGlyGluAlaSerValAspSerThrProTrpGlnValSerIle 220
 b 800 CCCCCTGTTGGTGGGAGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTTCAGCATC 859
 Y 221 GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
 b 860 CAGTACGACAAACAGACACGCTCTGTGGAGGAGCATCTGTGACCCCACTGGGTCTCAGC 919
 Y 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260
 b 920 GCAGCCCACTGCTTCAGGAAACATACCATGTGTTCAACTGGAAGGTGCGGCGAGCTCA 979
 Y 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsn 280
 b 980 GACAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATATTGAATTCAAC 1039
 Y 281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
 b 1040 CCCATGTACCCCAAGACAAATGACATCGCCTCATGAAGTGCAGTTCCTCACTCACTTC 1099
 Y 301 SerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThr 320

1100 TCAGGCACAGTCAGGCCCACTCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 1159
 QY 321 ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle 340
 DB 1160 CCACTCTGATCATTTGGATGGGCTTTACGAAGCAAGATGGAGGAAGATGTCTGACATA 1219
 QY 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyr 360
 DB 1220 CTGCTGCAGGCTCAGTCCAGTCAATGACAGCACAGGTGCAATGCAGACGATGCGTAC 1279
 QY 361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr 380
 DB 1280 CAGGGGGAAGTCACCGAGAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACC 1339
 QY 381 CysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGly 400
 DB 1340 TGCCAGGTTGACAGTGTGGGCCCTGATGTACCAATCTGACCATGTGGTGGGC 1399
 QY 401 IleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVal 420
 DB 1400 ATCGTTAGCTGGGCTATGGCTGGGGGGCCGAGCACCCAGGAGTATACACCAAGGTC 1459
 QY 421 SerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
 DB 1460 TCAGCCTATCTCACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1504

RESULT 4
 AF179224 2081 bp mRNA linear PRI 08-JUN-2000
 LOCUS Homo sapiens transmembrane serine protease 3 (TMPRSS3) mRNA,
 DEFINITION complete cds.
 ACCESSION AF179224
 VERSION AF179224.1 GI:8347148
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2081)
 AUTHORS Wallrapp,C., Hahnel,S., Muller-Pillasch,F., Burghardt,B.,
 Iwamura,T., Ruthenburger,M., Lerch,M.M., Adler,G. and Gress,T.M.
 TITLE A novel transmembrane serine protease (TMPRSS3) overexpressed in
 pancreatic cancer
 JOURNAL Cancer Res. 60 (10), 2602-2606 (2000)
 MEDLINE 20283276
 PUBMED 10825129
 REFERENCE 2 (bases 1 to 2081)
 AUTHORS Wallrapp,C. and Gress,T.M.
 TITLE Direct Submission
 JOURNAL Submitted (20-AUG-1999) Internal Medicine 1, University of Ulm,
 Robert-Koch-Street 8, Ulm, Baden-Wuerttemberg 89081, Germany

FEATURES
 source
 1. .2081
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11q23.3"
 /tissue_type="pancreatic carcinoma"
 1. .2081
 /gene="TMPRSS3"
 215. .1528
 /gene="TMPRSS3"
 /codon_start=1
 /product="transmembrane serine protease 3"
 /protein_id="AAF74526.1"
 /db_xref="GI:8347148"
 /translation="MLQDPDSDQPLNSLDVKPLRKPRIIPMETFRKVGIPIIALLSLA
 SIIIVVLKIVLDKYFLCGQLPHFIPRKQLCDGELDCPLGEDEEHCKVKSFEPPAV
 AVRISKDRSTLQVLDSATGNWFSACFDNFTALAEATACRQMGYSKPTFRVAVIGPDQ
 DLDVVEITENSQELRMNRSGLSLVSLHCLACGSKLTPRVVVGEEASVDSWPW
 QVSIQYDKQHVCGSILDPHWLTAHCFKHTDVFVNWKVRAGSKLGSFSPSLAVAKI
 IIEFNPMPYKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATPLWIIGWFTTKQ

```

NGKMSDILLQASVQVIDSTRCNADDAVQGEVTEKMMCMCAGIPBGGVDTCCQDSSGGLM
YQSDQWHVGVISWGYGGPSTPGVYTKVSAYLNWIYNVWKAEL"
215. .307
/misc_feature /gene="TMPRSS3"
/note="Region: cytoplasmic domain"
308. .373
/misc_feature /gene="TMPRSS3"
/note="Region: transmembrane domain"
374. .1525
/misc_feature /gene="TMPRSS3"
/note="Region: extracellular domain"
602. .604
/misc_feature /gene="TMPRSS3"
/note="glycosylation site"
746. .748
/misc_feature /gene="TMPRSS3"
/note="disulfide bond"
order(902. .904,950. .952)
/misc_feature /gene="TMPRSS3"
/note="disulfide bond"
947. .949
/misc_feature /gene="TMPRSS3"
/note="active site"
1082. .1084
/misc_feature /gene="TMPRSS3"
/note="active site"
order(1280. .1282,1328. .1330)
/misc_feature /gene="TMPRSS3"
/note="disulfide bond"
1373. .1375
/misc_feature /gene="TMPRSS3"
/note="active site"
484 a 597 c 576 g 424 t
BASE COUNT
REGION
Ligment Scores:
red. No.: 5.26e-214 Length: 2081
core: 2337.00 Matches: 434
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
Gaps: 0
us-09-607-745-2 (1-435) x AF179224 (1-2081)
2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
224 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAACCCCTGCGCAACCCCGT 283
22 IlePrometGluThrPheArgLysValGlyIleProIleIleIleAlaLeuSerLeu 41
284 ATCCCCATGGAGACCTTCAGAAAGGTGGGGATCCCCCATCATCATAGCACTACTGAGCCTG 343
42 AlaSerIleIleIleValValIleValIleLysValIleLysValIleLysValIle 61
344 GCGAGTATCATCATTTGTGTTGTCTCATCAAGGTGATCTCGATAAATACTACTTCTCCTC 403
62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
404 TCGGGGAGCCTCTCCACTTCATCCGAGGAGGAGCAGCTGTGTGACGGAGAGTGGACTGT 463
82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
464 CCCTTGGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCGCTGCAGTGGCA 523
102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
524 GTCCGCCTCTCCAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGGAAGTGG 583
122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141

```

```

Db 584 TTCTCTGCCTGTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCCTGTAGGCAGATG 643
QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
Db 644 GGCTACAGCAGCAACCCACTTTCAGAGCTGTGAGATTGGCCCGACAGCAGGATCTGGAT 703
QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
Db 704 GTTGTGAAATACAGAAACAGCCAGGAGCTTCGATGCGGAACTCAAGTGGGCCCTGT 763
QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
Db 764 CTCACAGCTCCCTGGTCTCCCTGCACCTGTCTGCTGTGGAGAGAGCTGAAGACCCCC 823
QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
Db 824 CGTGTGGTGGTGGGAGGAGGCTCTGTGATTTCTTGGCCTTGGCAGTCCAGCATCCAG 883
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
Db 884 TACGACAAACAGCACGCTCTGTGGAGGAGCATCTCGACCCCACTGGGTCTCTCAGGCA 943
QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
Db 944 GCCCACTGCTTCAGGAACATACCGATGTGTTCACAGTGAAGGTGCGGCGAGGCTCAGAC 1003
QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
Db 1004 AAACCTGGGAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATTAATTAACCCCC 1063
QY 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
Db 1064 ATGTACCCCAAGACAATGACATCGCCCTCATATGAGCTGCAGTTCCTCCACTCTTCTCA 1123
QY 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
Db 1124 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1183
QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
Db 1184 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGGAAGATGTCTGACATACTG 1243
QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 361
Db 1244 CTGCAGGCGTCAGTCCAGTCCATTCAGACACACAGCTGCAATGCAGACGATCGGTACCAG 1303
QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
Db 1304 GGGGAAGTCACCGAGAGATGATGTGTGAGGCAATCCCGAAGGGGTGTGGACACCTGC 1363
QY 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
Db 1364 CAGGGTGACAGTGTGGGCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGCATC 1423
QY 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
Db 1424 GTTAGCTGGGCTATGGCTGGGGGGGGCCGAGCAGCAGGAGTATACACCAAGTCTCA 1483
QY 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
Db 1484 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1525
RESULT 5
BC011703 2088 bp mRNA linear PRI 02-AUG-2001
LOCUS Homo sapiens, Similar to mosaic serine protease, clone MGC:19490
DEFINITION IMAGE:3610695, mRNA, complete cds.
ACCESSION BC011703
VERSION BC011703.1 GI:15079794
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2088)

REFERENCE

Strausberg, R.

AUTHORS

Direct Submission

TITLE

Submitted (30-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgaps-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@nhgri.nih.gov

Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,

Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,

Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,

Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,

Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,

Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>

Series: IRAL Plate: 27 Row: f Column: 19

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 8347148.

FEATURES

Location/Qualifiers

1..2088

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:19490 IMAGE:3610695"

/tissue type="Pancreas, adenocarcinoma"

/clone_lib="NIH_MGC_39"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

217..1530

/codon_start=1

/product="Similar to mosaic serine protease"

/protein_id="AAH11703.1"

/db_xref="GI:15079795"

/translation="MLQDFSDQPLNSLDVPLRKPRIPIPMETFRKVCIPILIALLSLA
SIIIVVLTKVILDKYFLCGPLHPIPRKQLDCGLDCLGEDEHCVKVSFPEGPVAV
AVRLSKDRSTLQVLDTSATGNWFSACFDNFTEALABTACRMGYSSKPTFRAVEIGPDQ
DLDVVEITNSQELRMNRSGLSLVSLHCLACGSKLKTPTRVVGGEEASVDSPWP
QVSIQYDKQHVCGGSLDPHWLTAHCFRKHDTFVNWKVRAGSKLGSFPPLAVAKI
IIIEFNMPYKDNIDIALMKLPPLTFSGTVRPICLPFFDEELTPATPLWIIIGWFTKQ
NGGKMSDILLQASVQVIDSTRNADDAYQGEVTERKMCAGIPEGGVDTFCQGDSSGGPLM
YQSDQWHVVGIVSWGCGGSPSTPGVYTKVSAVLNWIYNVWKAEL"

BASE COUNT

492 a

597 c 579 g 420 t

Ligment Scores:

red. No.: 5.28e-214 Length: 2088
core: 2337.00 Matches: 434
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
B: 9 Gaps: 0

S-09-607-745-2 (1-435) x BC011703 (1-2088)

2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21

|||||

226 GATCCTGACAGTGATCAACCTCTGTGACAGCTCGATGTCTCAAAACCCCTGCGCAAAACCCCGT 285

Qy 22 IleProMetCltThrPheArgLysValGlyIleProIleIleAlaLeuLeuSerLeu 41
|||||
Db 286 ATCCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATAGCAGCTACTGAGCCTG 345
Qy 42 AlaSerIleIleValValLeuIleLysValIleLeuAspLysTyrPheLeu 61
|||||
Db 346 GCGAGTATCATCATTTGGTTGCTCTCATCAAGGTGATTTCTGGATAAATACTACTTCTCTC 405
Qy 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
|||||
Db 406 TCGGGGCGAGCTCTCCACTTCATCCCGAGGAAGCAGTGTGTGACGGAGAGCTGGACTGT 465
Qy 82 ProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaValAla 101
|||||
Db 466 CCCTTGGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCCTGCAGTGGCA 525
Qy 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
|||||
Db 526 GTCCGCTCTCCAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGAACCTGG 585
Qy 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
|||||
Db 586 TTCTCTGCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAGATG 645
Qy 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
|||||
Db 646 GGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCAGACAGGATCTGGAT 705
Qy 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
|||||
Db 706 GTTGTGTAATACAGAAACACAGGAGCTTCGATGCGGAACCTCAAGTGGGCCCTGT 765
Qy 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
|||||
Db 766 CTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCCTGTGGAGAGCCCTGAAGACCCCC 825
Qy 202 ArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
|||||
Db 826 CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTGAGCATCCAG 885
Qy 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
|||||
Db 886 TACGACAAACAGCACACGTCTGTGAGGGAGCATCTCTGACCCCTTGGGTCTCTCACGCA 945
Qy 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
|||||
Db 946 GCCCACTGTTCAGGAACATACCGATGTGTTCACTGGAAGGTGCGGGCAGGCTCAGAC 1005
Qy 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
|||||
Db 1006 AAACCTGGGCGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATATTGAATCAACCCC 1065
Qy 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
|||||
Db 1066 ATGTACCCCAAGACATGACATGACATGCGCCCTCATGAAGTGCAGTTCCCACTCACTTTCTCA 1125
Qy 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
|||||
Db 1126 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1185
Qy 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
|||||
Db 1186 CTCTGGATCATTTGGATGGGGCTTTACGAAGCAGATGAGGGAAGATGTCTGACATACTG 1245
Qy 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 361
|||||
Db 1246 CTGCAGGCGTCAGTCCAGGTTCATTGACAGCACACCGTGCATGCAGACGATGCTGATCCAG 1305
Qy 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
|||||
Db 1306 GGGGAAGTCACCGAGAAGATGATGTGTGAGGATATCCCGAAGGGGTGTGACACCTGC 1365
Qy 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401


```
|||||
1366 CAGGGTGACAGTGGTGGCCCTGATGTACCAATCTGACAGTGGCATGTGGTGGCATC 1425
|||||
402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
|||||
1426 GTTAGCTGGGCTATGGCTGGGGGCGCGAGCACCCAGGAGTATACACCAAGGTCTCA 1485
|||||
422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
|||||
1486 GCCTATCTCACTGAGTCTACAAATGCTTGAAGGCTGAGCTG 1527
|||||

RESULT 6
AX076192          1479 bp      DNA      linear      PAT 06-FEB-2001
LOCUS
DEFINITION      Sequence 3 from Patent WO0104141.
ACCESSION      AX076192
VERSION
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS      Suendermann,B., Hofmann,U., Matzku,S. and Wilbert,O.
TITLE      Seripancrin
JOURNAL      Patent: WO 0104141-A 3 18-JAN-2001;
MERCK PATENT GmbH (DE)
FEATURES
source
1..1479
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1..1479
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC28480.1"
/db_xref="GI:12710818"
/translation="MDPDSQDPLNSLDVKPLRPIMPETFRKVGIPPIIIALLSLASI
IIVVLIKVILDKYFLCGQLHFI PRKQLCDGELDCPLGEDEHCVKSFPEGPAAV
RLSKDRSTLQVLD SATGNWFSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDL
DVVEITENSQELRMNRSSGCLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQV
SIQYDKHVCVGGSI LDPHWLTAARCFRKHDTDFNWKVRAGSDKLGSPPSLAVAKIII
IEFNPMYKDNIDIALMKLQFLPTFSGTVRPICI PFDEELTPATPLWIGWFTKQNG
GKMSDILLQASVQVIDSTRCNADDAVQGEVTEKMMKAGIPEGGVDTCCQSDSGGLMYQ
SDQWHVVGIVSWGYCGGSPSTPGVYTKVSAYLNIYNWVKDRTIQRCNSPSTGLVIG
QPAVPLMELRSARHTLRLTTCVCFLSFPAQLPLSPFPFPWD"
ASE COUNT      329 a 425 c 407 g 318 t
RIGIN

Alignment Scores:
red. No.:      1.97e-213      Length:      1479
core:          2329.00      Matches:      432
percent Similarity: 100.00%      Conservative: 0
est Local Similarity: 100.00%      Mismatches: 0
query Match:    99.44%      Indels:      0
B:             6      Gaps:      0

S-09-607-745-2 (1-435) x AX076192 (1-1479)

Y      1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
b      1 ATGGATCCTGACAGTGTCAACCTCTGAACAGCCTCGATGTCAACCCCTGGCGAAACCC 60
Y      21 ArgIleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSer 40
b      61 CGTATCCCCATGGAGACCTTCAGAAAGGTGGGGATCCCCATCATAGCACTACTGAGC 120
Y      41 LeuAlaSerIleIleIleValValLeuIleLysValIleLysValTyrTyrPhe 60
b      121 CTGGCGAGTATCATATTGTTGTTCTCATCAAGGTGATTCTGGATAAATACTACTTC 180
Y      61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
|||||
```

```
Db      181 CTCTGGGGCGACCTCTCCACTTCCATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGAC 240
QY      81 CysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaVal 100
Db      241 TGTCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCTGACAGTG 300
QY      101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
Db      301 GCAGTCCGCTCTCCAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACACGGGAAC 360
QY      121 TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
Db      361 TGGTTCTCTGCTGCTGTTTCGACAACTTCAGAAAGCTCTCGCTGAGACAGCCTGTAGGAC 420
QY      141 MetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
Db      421 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCCCAGACCATGTCG 480
QY      161 AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180
Db      481 GATGTTGTTGAATCACAGAAACACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCC 540
QY      181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
Db      541 TGTCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTGTGATTCTTGGCTTGGCAGGTGAGACC 600
QY      201 ProArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
Db      601 CCCGCTGTGGTGGGGAGGAGGCTCTGTGATTCTTGGCTTGGCAGGTGAGCATC 660
QY      221 GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
Db      661 CAGTACGACAAACAGCACGCTCTGTGGAGGAGCATCTCTGACCCCTGCTGGCTCCTCAG 720
QY      241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260
Db      721 GCAGCCCACTGTTTCAAGAAACATACCGATGTGTTCACCTGGAAGGTGCGGCGAGCTCA 780
QY      261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsn 280
Db      781 GACAAACTGGGAGCTTCCCATCTCCCTGGCTGTGGCCAAAGATCATCATCATTTGAATTCAAC 840
QY      281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
Db      841 CCCATGTACCCCAAAGACAAATGACATCGCCCTCATGAAGCTGAGTCCCACCTCACTTTC 900
QY      301 SerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThr 320
Db      901 TCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGTCACTCCAGCCACC 960
QY      321 ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIle 340
Db      961 CCACTCTGGATCATGTGGATGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATA 1020
QY      341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyr 360
Db      1021 CTGCTGCAGGCTCAGTCCAGGTCAATGACAGCACACCGTGTCAATGCAGACGATGCGTAC 1080
QY      361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr 380
Db      1081 CAGGGGGAAGTCACCGAGAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACC 1140
QY      381 CysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGly 400
Db      1141 TGCCAGGTGACAGTGTGGGGCCCTGATGTACCAATCTGACCATGTGGCATGTGGTGGGC 1200
QY      401 IleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVal 420
Db      1201 ATCGTTAGCTGGGCTATGGCTGGGGGGCCCGAGCACCCGAGGAGTATACACCAAGGTC 1260
QY      421 SerAlaTyrLeuAsnTrpIleTyrAsnValTrpLys 432
Db      1261 TCAGCCTATCTCAACTGGATCTACAAATGTCTGGAAG 1296
```


Db	817	CCGTGTGGTGGTGGGAGGAGCCCTCTGTGGATTCTTGGCCTTGGCAGGTACAGCATCCA	876
Qy	221	nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl	241
Db	877	GTACGACAAACAGCACGTCTGTGGAGGAGCATCCTGGACCCCTACTGGGTCTTCACGGC	936
Qy	241	aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs	261
Db	937	AGCCACTGCTTTCAGGAAACATACCGATGTGTTCAACTGGAAAGGTGCGGCGAGGTCTAGA	996
Qy	261	pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPr	281
Db	997	CAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATCATTTCTC	1056
Qy	281	oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe	301
Db	1057	CATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAAGTCCCACTCACTTTCTC	1116
Qy	301	rGlyThrValArgProLysCysLeuProPhePheAspGluGluLeuThrProAlaThrPr	321
Db	1117	AGGCACAGTCAGGCCCATCTGTCTGCCCTCTTTTGATGAGGAGCTCACTCCAGCCACCC	1176
Qy	321	oLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLe	341
Db	1177	ACTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTTGACATACT	1236
Qy	341	uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGl	361
Db	1237	GCTGCAGCGTCAGTCAGGTCATTTGACAGCACACGGTGCATGTCAGACGATGCGTACCA	1296
Qy	361	nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy	381
Db	1297	GGGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGGAAGGGGTGTGGACACCTG	1356
Qy	381	sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyI	401
Db	1357	CCAGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAAGTGCATGTGGTGGGCAT	1416
Qy	401	eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe	421
Db	1417	CGTTAGCTGGGGCTATGGCTCGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTC	1476
Qy	421	rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	435
Db	1477	AGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG	1519
RESULT	8		
AX207967			
LOCUS	AX207967	2079 bp	DNA linear PAT 31-AUG-2001
DEFINITION	Sequence 71 from Patent WO0157194.		
ACCESSION	AX207967		
VERSION	AX207967.1	GI:15422563	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1		
AUTHORS	Madison, E.L., Ong, E.O. and Yeh, J.C.		
TITLE	Nucleic acid molecules encoding transmembrane serine proteases, the encoded proteins and methods based thereon		
JOURNAL	Patent: WO 0157194-A 71 09-AUG-2001;		
FEATURES	CORVAS INTERNATIONAL, INC. (US)		
source	Location/Qualifiers		
	1. .2079		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
CDS	251..1522		
	/note="Nucleotide sequence encoding transmembrane serine protease, serine 4 (TPRSS4)"		
	/codon_start=1		

/protein_id="CAC60389.1"
/db_xref="GI:15422564"
/translation="MSNPANPVPSPRPSPESVGIPIIIALLSLASIIIVVLIKVILD
KYFLCGQPLHFIPIKQDCGELDCPLGEDEHCVKSPFEGPAVAVRLSKDRSTLOVL
DSATGNWFSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQEL
RMNRSGLCLSGSLHCLACGSKLKTPTVVVGEEASVDSWPVOVSIQYDKQHVCGG
SILDPHWLTAACHCRKHTDVFNWVKVRAGSKLSPPLSAVAKILIIIEPNMYPKDN
IALMKLQPLFTFSGTVPRIICLPFFDEELTPATPLMIIGWFTKONGKMSDILLQASV
QVIDTRCNADDAAYQGEVTEKMKCAGIPEGGVDTCCQSDSGGPLEMYQSDQWHVVGIVSW
GYCGGPSTPGVYTKVSAYLNIWVWKAEL"
BASE COUNT 489 a 594 c 575 g 421 t
ORIGIN

Alignment Scores:
pred. No.: 9,26e-213 Length: 2079
core: 2324.00 Matches: 434
percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.23% Indels: 1
Gaps: 0

US-09-607-745-2 (1-435) x AX207967 (1-2079)

Qy	2	AspProAspSerAspGlnProLeuAsnSerLeuAspVallysProLeuArgLysProArg	21
Db	217	GATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGGCGAAACCCCGT	276
Qy	22	IleProMetGluThrPheArgLys-ValGlyIleProIleIleIleAlaLeuLeuSerLe	41
Db	277	ATCCCCATGGAGACCTTCAGAAAGTGGGGATCCCCATCATATAGCACTACTGAGCCT	336
Qy	41	uAlaSerIleIleIleValValleulleLysValIleLeuAspLysTyrTyrPheLe	61
b	337	GGCGAGTATCATCATGTGTGTTGCTCATCAAGGTGATTCGGATAAATACTACTTCCT	396
Y	61	uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCy	81
b	397	CTGCGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGA	456
Y	81	sProLeuGlyGluAspGluGluHisCysVallysSerPheProGluGlyProAlaValAl	101
b	457	TCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTGAGTGGC	516
Y	101	aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr	121
b	517	AGTCCGCCTCTCCAGGACCGATCCACACTGCAGGTGTGTGACTCGGCCACAGGAACTG	576
Y	121	pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe	141
b	577	GTCTCTGCTGCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAG	636
Y	141	tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs	161
b	637	GGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCAGACCAGGATCTGA	696
Y	161	pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy	181
b	697	TGTTGTTGAATCACAGAAACACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGCCCTG	756
Y	181	sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr	201
b	757	TCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGGCTGTGGGAAGAGCCTGAAGACCCC	816
Y	201	oArgValValGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleG1	221
b	817	CCGTGTGGTGGGAGGAGGAGGAGGCTCTGTGGATTCTTGGCCTGGCAGGTGAGATCCA	876
Y	221	nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl	241
b	877	GTACGACAAACAGCAGCTCTGTGGAGGAGCATCTTGGACCCCACTGGTCTCTCAGCGC	936
Y	241	aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs	261

Db	937	AGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCAGA	996
Qy	261	pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPr	281
Db	997	CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATGATTGATCAACCC	1056
Qy	281	oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe	301
Db	1057	CATGTACCCCAAGACAATGACATGCCCTCATGAAGCTGCAGTTCACCACTCACTTTCTC	1116
Qy	301	rGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPr	321
Db	1117	AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCC	1176
Qy	321	oLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLe	341
Db	1177	ACTCTGGATCATTTGGATGGGCTTTACGAAGCAGATCGAGGAAGATGTCTGACATACT	1236
Qy	341	uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrG1	361
Db	1237	GCTCAGGCGTCAGTCCAGTCAATGACAGCACACCGTCAATGCAGACGATGCGTACCA	1296
Qy	361	nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy	381
Db	1297	GGGGAAAGTCACCGAGAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGACACCTG	1356
Qy	381	sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyI1	401
Db	1357	CCAGGTGACAGTGGTGGCCCTCATGTACCAATCTGACCAAGTGGCATGTGTGGGCGAT	1416
Qy	401	eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValse	421
Db	1417	CGTTAGCTGGGCTATGGCTGCGGGGCCCCGAGCACCCAGGAGTATACACCAAGCTCTC	1476
Qy	421	rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	435
Db	1477	AGCCTATCTCACTGGATCTACATGTCTGGAAGGTGAGCTG	1519

RESULT 9
AF216312
LOCUS AF216312 2079 bp mRNA linear PRI 07-FEB-2000
DEFINITION Homo sapiens type II membrane serine protease mRNA, complete cds.
ACCESSION AF216312
VERSION AF216312.1 GI:6911218
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2079)
AUTHORS Smeekens, S.S., Lorimer, D.D., Wang, E., Hou, J. and Linnevers, C.
TITLE MT-SP2, a novel type II membrane serine protease expressed in trachea, colon, and small intestine: identification, cloning, and chromosomal localization
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2079)
AUTHORS Smeekens, S.S., Lorimer, D.D., Wang, E., Hou, J. and Linnevers, C.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-1999) Axys Pharmaceuticals, Inc, 180 Kimball Way, South San Francisco, CA 94080, USA
FEATURES
Location/Qualifiers
1. .2079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
251. .1522
/note="MT-SP2"
/codon_start=1
/product="type II membrane serine protease"
/protein_id="AAF31436.1"
/db_xref="GI:6911219"
/translation="MSNPANPVPSPRPSPESVGIPIIIALLSLASIIIVVLIKVILD
KYFLCGQPLHFIPIKQDCGELDCPLGEDEHCVKSPFEGPAVAVRLSKDRSTLQVL

DSATGNWFSACFDNFTEALAEATACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQEL
RMRNSSPCLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPWQVSIQYDKQHVCGG
SILDPHWLTAAGHGRKHTDVFNWKRAGSDKLGSPSLAVAKIIIEFNPMYKDNND
IALMKLQPLTFSGTRVPCILPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASV
QVIDSTRNADDAIQGEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQWHVVGIVS
GYCGGPGSTPGVYTKVSAYLNIWNVKAEI"
BASE COUNT 489 a 594 c 575 g 421 t
ORIGIN

Alignment Scores:
Pred. No.: 9.26e-213 Length: 2079
Score: 2324.00 Matches: 434
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.23% Indels: 1
DB: 9 Gaps: 0

US-09-607-745-2 (1-435) x AF216312 (1-2079)

2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
217 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGAAACCCCGT 276
22 IleProMetGluThrPheArgLys-ValGlyIleProIleIleIleAlaLeuLeuSerLe 41
277 ATCCCATGGAGACCTTCAGAAAGTGTGGGATCCCATCATCATAGCACTACTGAGCCT 336
41 uAlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrPheLe 61
337 GCGAGTATCATATTGTTGTCTCATCAAGGTGATTCCTGGATAAATACTACTTCTT 396
61 uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCy 81
397 CTGCGGGCAGCCTCTCCACTTCTATCCCGAGGAGCAGCTGTGTACGGAGAGCTG 456
81 sProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAl 101
457 TCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTGCAGTGC 516
101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
517 AGTCCGCTCTCCAAGGACCGATCCACACTGAGGTGTGCTGAGCTCGGCCACAGGGA 576
121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
577 GTTCTCTGCCTGTTTCACAACTTCACAGAACTCTCGCTGAGACAGCCTGTAGGAGAT 636
141 tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs 161
637 GGGTACAGCAGCAACCACTTTCAGAGCTGTGGAGATTGGCCAGACCCAGGATCTGGA 696
161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181
697 TGTGTGTAATCACAGAAACAGCCAGGAGCTTGCATGCGGAATCAAGTGGGCCCTG 756
181 sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr 201
757 TCTCTCAGGCTCCCTGTTCTCCCTGCATGCTTGCCTGTGGAGAGCCTGAAGACCCC 816
201 oArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIleGl 221
817 CCGTGTGGTGGTGGGAGGAGGCTCTGTGGATCTTGGCCTTGGCAGGTCAAGATCCA 876
221 nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl 241
877 GTACGACAAACAGCAGCTCTGTGGAGGAGGATCCTGGACCCCTTGGGTCTCTACGGC 936
241 aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261
937 AGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGTGGGCGGAGGCTCAGA 996
261 pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPr 281

Db 997 CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCACAGATCATCATTTGAATCAACCC 1056
Qy 281 oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301
1057 CATGTACCCCAAGACAAATGACATCGCCCTCATGAAGTGCAGTTCCTCACTACTTCTC 1116
Qy 301 rGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPr 321
1117 AGGCACAGTCAGGCCCATCTGTCTGCTCTCTTTGATGAGGAGCTCACTCCAGCCACCCC 1176
Qy 321 oLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLe 341
1177 ACTCTGGATCATTTGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACT 1236
Qy 341 uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGl 361
1237 GCTGCAGGGCTCAGTCCAGGTCAATGACAGCACACCGTCAATGCAGACGATGCGTACCA 1296
Qy 361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381
1297 GGGGAAGTCAACGAGAGATGATGTGTGCAGGCATCCCGAAGGGGTGTGGACACCTG 1356
Qy 381 sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIl 401
1357 CCAGGTGACAGTGGTGGGCCCTCATGTACCAATCTGACAGTGGCATGTGTGGGCGAT 1416
Qy 401 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe 421
1417 CGTTAGCTGGGCTATGGCTGCGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTC 1476
Qy 421 rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
1477 AGCCTATCTCACTGGATCTACATGTCTGGAAGGCTGAGCTG 1519
Db

RESULT 10

AX207899 LOCUS 2137 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 3 from Patent WO0157194.
ACCESSION AX207899
VERSION AX207899.1 GI:15422497
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Madison, E.L., Ong, E.O. and Yeh, J.C.
TITLE Nucleic acid molecules encoding transmembrane serine proteases, the
encoded proteins and methods based thereon
JOURNAL Patent: WO 0157194-A 3 09-AUG-2001;
CORVAS INTERNATIONAL, INC. (US)

FEATURES

source
1..2137
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
261..1574
/note="DNA sequence encoding a transmembrane serine
protease (MTSP3) protein"
/codon_start=1
/protein_id="CAC60378.1"
/db_xref="GI:15422498"
/translation="MLQDPDSQPLNSLDVPLKPRIPMETFRKVGIPPIIALLSLA
SIIIVVLIKVIDKYIFLCQQLHFIPIKQLCDGELDCPLGEDEHCVKSPFEGPAV
AVRLSKDRSTLQVLDATGNWFSACFDNFTEALAEATACRQMGYSKPTFRAVEIGPDQ
DLDVVEITENSQELMRNSSPCLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPW
QVSIQYDIQHVCGGSIILDPHWLTAHCFRKHDPVFNWKRAGSDKLGSPPSLAVAKI
IIEFNPMYKDNNDIALMKLQPLTFSGTRVPCILPFFDEELTPATPLMIIGWFTKQ
NGKMSDILLQASVOVIDSTRNADDAIQGEVTERKMCAGIPEGGVDTCCQDSGGPLM
YQSDQWHVVGIVSWGYGCGGPGSTPGVYTKVSAYLNIWNVKAEI"

CDS
BASE COUNT 508 a 612 c 589 g 428 t
ORIGIN

Alignment Scores:

core:	2.89e-212	Length:	2137
Percent Similarity:	2319.00	Matches:	432
Best Local Similarity:	99.54%	Conservative:	0
Query Match:	99.54%	Mismatches:	2
DB:	99.02%	Indels:	0
	6	Gaps:	0

JS-09-607-745-2 (1-435) x AX207899 (1-2137)

2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
270 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCTAAACCCCTGGCGCAACCCCGT 329

22 IleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuSerLeu 41
330 ATCCCCATGGAGACCTTCAGAAAGTGGGATCCCCATCATCATAGCACTACTAGCCTG 389

42 AlaSerIlellelleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
390 GCGAGTATCATCATTTGGTTGTCTCATCAAGGTGATTCGGATAAATACTACTTCCTC 449

62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGlnLeuAspCys 81
450 TGCGGGCAGCCTTCACCTTCATCCCGAGGAAGCAGCTGTGTACCGAGAGCTGGACTGT 509

82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProLaValAla 101
 510 CCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCCGTCAGTGGCA 569

102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
|||||
570 GTCCGCCTCTCCAAGGACCGATCCACACTGCAGTGCTGGACTCGGCCACAGGAACTGG 629

122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
630 TTCTCTGCCTGTTTCGACAACTTCACAGAAAGCTCTCGCTGAGACAGCCCTTAGGCAGATG 689

y 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
 |||||
 b 690 GGCTACAGCAGCAAPACCCACCTTCAGAGCTGTGGAGATTGGCCACACCGATCTGGAT 749

y 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
 |||||
 b 750 GTTGTTGAAATCACAGAAAAACAGCCAGGAGCTTCGCATCGGGAACCTCAACTGGGCCCCGT 809

y 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
 |||||
 b 810 CTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTGGCCTGTGGGAAGAGCGCTGAAGACCCCC 869

202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
 870 CGTGTGTGGTGGGGAGGAGGCCCTCTGTGGATTCTTGGCCCTTGGCAGTCCAGTCAGATCCAG 929

222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
 930 TACGACATACAGCACGCTCTGTGGAGGGAGCATCTGGACCCCACTGGGTCTCTACGGCA 989

242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
 |||||
 990 GCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGCAGGCTCAGAC 1041
 |||||

y 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
 |||||
 b 1050 AAATGGGAGCTTCCCATCCCTGGCTGTGGCCAAGATCATCATCATGAATTCAACCCC 1109

282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
1110 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCCTCACTTCTCA 1169

302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
 1170 GGACAGTCAGGCTCATCTGTCTGCGCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1229

QY	322	Leu	Trp	Ile	Ile	Gly	Trp	Gly	Phe	Thr	Lys	Gln	Asn	Gly	Gly	Lys	Met	Ser	Asp	Ile	Leu	341
DB	1230	CT	CT	GG	AT	CA	TG	AT	GG	GC	T	T	AC	GA	AG	CA	GA	AT	GG	AG	GA	1289
QY	342	Leu	Gln	Ala	Ser	Val	Gln	Val	Ile	Asp	Ser	Thr	Arg	Cys	Asn	Ala	Asp	Ala	Tyr	Gln	361	
DB	1290	CT	GC	AG	CG	CT	CAG	GT	CA	T	T	GA	CA	CA	CG	GT	CA	AT	GC	AG	1349	
QY	362	Gly	Glu	Val	Thr	Glu	Lys	Met	Met	Cys	Ala	Gly	Ile	Pro	Glu	Gly	Gly	Val	Asp	Thr	Cys	381
DB	1350	GG	GA	AG	TC	AC	CG	AG	AG	AT	GA	T	GT	GT	CA	GG	CA	T	CC	CG	1409	
QY	382	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Met	Tyr	Gln	Ser	Asp	Gln	Trp	His	Val	Val	Gly	Ile	401
DB	1410	CAG	GT	GA	CA	GT	GT	GG	CC	CC	CT	GA	T	GA	T	AC	CA	AT	CT	GA	1469	
QY	402	Val	Ser	Trp	Gly	Tyr	Gly	Cys	Gly	Pro	Ser	Thr	Pro	Gly	Val	Tyr	Thr	Lys	Val	Ser	421	
DB	1470	GT	TA	GT	GG	GG	CT	AT	GG	CT	CG	GG	CC	CG	AG	CA	CC	CC	AG	AG	1529	
QY	422	Ala	Tyr	Leu	Asn	Trp	Ile	Tyr	Asn	Val	Trp	Lys	Ala	Glu	Leu						435	
DB	1530	GC	CT	AT	CT	CA	ACT	GG	AT	CT	CA	AT	GT	CT	GG	AG	GG	CT	GAG	CTG	1571	

RESULT 11

AX092380	AX092380	2063 bp	DNA	linear	PAT 21-MAR-2001
LOCUS	Sequence 111	from Patent WO0116318.			
DEFINITION	AX092380				
ACCESSION	AX092380.1	GI:13444504			
VERSION					

KEYWORDS
SOURCE
Homo sapiens (human)

ORGANISM	Homo sapiens	Homo sapiens (modern)
Source		

REFERENCE

AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1

Eaton, D.L., Filvaroff, E., Gerritsen, M.E., Goddard, A., Godowski, P.J., Grimaldi, C.J., Gurney, A.L., Watanabe, C.K. and Wood, W.I.

TITLE

JOURNAL Patent: WO 0116318-A 111 08-MAR-2001;
Genentech, Inc. (US)

FEATURES

source
1. .2063

```
/organism="Homo sapiens"  
/mol_type="genomic DNA"  
/db_xref="taxon:9606"
```

	477 a	591 c	576 g	419 t
BASE COUNT				
ORIGIN				

Alignment Scores:

Pred. No.:	3.17e-210	Length:	2063
Score:	2297.50	Matches:	429
Percent Similarity:	98.85%	Conservative:	0
Best Local Similarity:	98.85%	Mismatches:	0
Query Match:	98.10%	Indels:	5
DB:	6	Gaps:	1

US-09-607-745-2 (1-435) x AX092380 (1-2063)

Qy		2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg	21
D6		219 GATCTTGACAGTGTCAACTCTGAACAGCCTCGATGTCAAACCCTGCGCCAACCCCTG	278

Qy		22	IleProMetGluThrPheArgLysValGlyIleLeProIleIleIleAlaLeuLeuSerLeu	41
Dd		279	ATCCCCATGGAGACCTTCAGAAAGGTGGGGATCCCATCATCATAGCACTACTGAGCCTG	338

Qy 42 AlaSerIleIleValValValValValValLysValIleLeuAspLysTyrTyrPheLeu 61
Dd 339 CGCAGTATCATCATTTGGTGTCTCATCAAGGTGATCTGGATAAATACTACTCTCTC 398

QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysaspGlyGluLeuAspCys 81
Db 399 TCGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGACTGT 458
QY 82 ProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaValAla 101
Db 459 CCCTTGGGGAGGAGGAGGAGCAGCTGTGTCAAGAGCTTCCCCGAAGGGCCTGCAGTGGCA 518
QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
Db 519 GTCCGCCCTCTCCAGGACCGATCCACACTGCAGGTCTGGACTCGGCCACAGGGAAGTGG 578
QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
Db 579 TTCTCTGCCTGTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCCTGTAGGCAGATG 638
QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
Db 639 GGCTACAGC-----AGAGCTGTGGAGATTGGCCACAGACAGGATCTGGAT 683
QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
Db 684 GTTGTGAATCACAGAAAAACAGCCAGGAGTTCGCATGCGGAATCAAGTGGGCCCTGT 743
QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
Db 744 CTCTCAGGCTCCCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 803
QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
Db 804 CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGTTCAGCATCCAG 863
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
Db 864 TACGACAAACAGCAGCTGTGTGGAGGAGCATCTGACCCCTGCTGGTCTCTCAGGCA 923
QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValAlaGlySerAsp 261
Db 924 GCCCACTGCTTCAGGAACATACCGATGTGTTCACACTGGAAGTGGGGCAGGCTCAGAC 983
QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
Db 984 AACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATCATCATCATCATCATCAT 1043
QY 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
Db 1044 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCACACTCATCTTCTCA 1103
QY 302 GlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThrPro 321
Db 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCCCA 1163
QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
Db 1164 CTCTGGATCATTTGGATGGGCTTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATACTG 1223
QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 361
Db 1224 CTGAGGCGCTCAGTCCAGGTCAATTGACAGCACACGGTGCATGCGACAGCATGCGTACAG 1283
QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
Db 1284 GGGGAAGTCAACCGAGAGATGATGTGTGAGGATCCCGGATCCCGAAGGGGTGTGACACCTGC 1343
QY 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
Db 1344 CAGGGTGACAGTGGTGGGCCCTGTATGTACCAATCTGACCAAGTGGCATGTGGTGGGCATC 1403
QY 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
Db 1404 GTTAGCTGGGGCTATGGCTGG 1463

QY 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
Db 1464 GCCTATCTCACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505
RESULT 12
AX376262 2063 bp DNA linear PAT 01-MAR-2002
LOCUS AX376262
DEFINITION Sequence 329 from Patent WO0168848.
ACCESSION AX376262
VERSION AX376262.1 GI:19170522
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Baker, K.P., Chen, J., Desnoyers, L., Goddard, A., Godowski, P.J.,
Gurney, A.L., Pan, J., Smith, V., Watanabe, C.K., Wood, W.I. and
Zhang, Z.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0168848-A 329 20-SEP-2001;
Genentech, Inc. (US)
FEATURES
Location/Qualifiers
source
1..2063
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 477 a 591 c 576 g 419 t
ORIGIN
Alignment Scores:
Pred. No.: 3.17e-210 Length: 2063
Score: 2297.50 Matches: 429
Percent Similarity: 98.85% Conservative: 0
Best Local Similarity: 98.85% Mismatches: 0
Query Match: 98.10% Indels: 5
DB: 6 Gaps: 1
US-09-607-745-2 (1-435) x AX376262 (1-2063)
QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
Db 219 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAACCCCTGCGCAACCCCGT 278
QY 22 IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeu 41
Db 279 ATCCCATGGAGACCTTCAGAAAGGTGGGATCCCCATCATATAGCACTACTGAGCCTG 338
QY 42 AlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
Db 339 GCGAGTATCATCATTTGGTTGTCTTCATCAAGGTGATTTCTGGATAAATACTACTTCTCTC 398
QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
Db 399 TCGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGACTGT 458
QY 82 ProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaValAla 101
Db 459 CCCTTGGGGAGGAGGAGGAGCAGCTGTGTCAAGAGCTTCCCCGAAGGGCCTGCAGTGGCA 518
QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
Db 519 GTCCGCCCTCTCCAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGGAAGTGG 578
QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
Db 579 TTCTCTGCCTGTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCCTGTAGGCAGATG 638
QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
Db 639 GGCTACAGC-----AGAGCTGTGGAGATTGGCCACAGACAGGATCTGGAT 683

162 ValValGluLeuThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
|
684 GTTGTGAATCACAGAAACAGCCAGAGCTTCGCATGCGGAACCTCAAGTGGGCCCTGT 743
|
182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
|
744 CTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGCTGCTGTGGAGAGCCTGAAGACCCCC 803
|
202 ArgValValGlyGlyGluAlaSerValAspSerTrpTrpGlnValSerIleGln 221
|
804 CGTGTGGTGGGTGGGAGGAGGCCCTCTGTGGATTCTTGGCCTTGGCAGGTACGATCCAG 863
|
222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
|
864 TACGACAAACAGCACAGTCTGTGGAGGAGCATCTGGACCCCACTGGGTCTCTCACGGCA 923
|
242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
|
924 GCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGGTGCGGCGAGGTCCAGAC 983
|
262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
|
984 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATCTTGAATCAACCCC 1043
|
282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
|
1044 ATGTACCCCAAGACATACATCGCCCTCATGAAGCTGCAGTTCACACTCACTTCTCA 1103
|
302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
|
1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163
|
322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu 341
|
1164 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACTG 1223
|
342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 361
|
1224 CTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCAATGCAGACGATCGGTACCAG 1283
|
362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
|
1284 GGGGAAGTCACCGAGAAGATGATGTGTGAGGCATCCCGGAAGGGGTGTGGACACCTGC 1343
|
382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
|
1344 CAGGTTGACAGTGTGGGCCCTGATGTACCAATCTGACCACTGACAGTGGCATGTGGTGGCATC 1403
|
402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
|
1404 GTTAGCTGGGCTATGGCTGCGGGGGCCCGAGCACCCAGGAGTATACCAAGGTCTCA 1463
|
422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
|
1464 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505
|

RESULT 13
X395214
OCUS AX395214 2063 bp DNA linear PAT 18-MAY-2002
EFINITION Sequence 2 from Patent WO0216429.
CCESION AX395214
ERSION AX395214.1 GI:21066245
EYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1
Goddard, A., Godowski, P.J., Gurney, A.L., Hillan, K.J., Polakis, P.,
Smith, V., Wood, W.I., Wu, T.D. and Zhang, Z.
Compositions and methods for the diagnosis and treatment of tumor
Patent: WO 0216429-A 2 28-FEB-2002;
Genentech, Inc. (US)

FEATURES
source
Location/Qualifiers
1..2063
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 477 a 576 g 419 t
ORIGIN
Alignment Scores:
Pred. No.: 3.17e-210 Length: 2063
Score: 2297.50 Matches: 429
Percent Similarity: 98.85% Conservative: 0
Best Local Similarity: 98.85% Mismatches: 0
Query Match: 98.10% Indels: 5
DB: 6
US-09-607-745-2 (1-435) x AX395214 (1-2063)
QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
|
Db 219 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAACCCCTGCGCAAAACCCCGT 278
|
QY 22 IleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSerLeu 41
|
Db 279 ATCCCATGGAGACCTTCAGAAAGGTGGGGATCCCATCATCATAGCACTACTGAGCCTG 338
|
QY 42 AlaSerIleIleValValValLeuIleLysValIleLeuAspLysTyrPheLeu 61
|
Db 339 GCGAGTATCATCATTTGGTTGTCTCATCAAGGTGATTCTGGATAAATACTACTTCTCTC 398
|
QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
|
Db 399 TCGGGGCAGCCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGGAGAGCTGGACTGT 458
|
QY 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
|
Db 459 CCCTTGGGGAGGAGGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGCGCTGCAGTGGCA 518
|
QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
|
Db 519 GTCCGCTCTCCAGGACCGATCCACACTGCAGGTGTGACTCGGCACAGGAACTGG 578
|
QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
|
Db 579 TTCTCTGCCTGTTTCGACAACTTCACAGAAGCTCTCGTGAGACAGCTGTAGGCAGATG 638
|
QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
|
Db 639 GGCTACAGC-----AGAGCTGTGGAGATTGGCCCAACACAGGATCTGGAT 683
|
QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
|
Db 684 GTTGTGAATCACAGAAACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCCCTGT 743
|
QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
|
Db 744 CTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGCCTGTGGGAAGAGCCTGAAGACCCCC 803
|
QY 202 ArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
|
Db 804 CGTGTGGTGGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTTCAGCATCCAG 863
|
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
|
Db 864 TACGACAAACAGCACAGTCTGTGGAGGAGCATCTCTGGACCCCACTGGGTCTCTCACGGCA 923
|
QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
|
Db 924 GCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGGTGCGGCGAGGTCTCAGAC 983
|
QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
|
Db 984 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATCTTGAATCAACCCCC 1043
|

y	282	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	301
b	1044	ATGTACCCCAAGACATGACATGCCCTCATGAAGTGCAGTCCCACTCTCTCTCA	1103
y	302	GlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAlaThrPro	321
b	1104	GGCAGAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA	1163
y	322	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	341
b	1164	CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACTG	1223
y	342	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	361
b	1224	CTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGTCGAATGCAGACGATCGTACCG	1283
y	362	GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThrCys	381
b	1284	GGGGAAGTCACCGAGAAGATGATGTGCAGGCATCCCGAAGGGGTGTGGACACCTGC	1343
y	382	GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	401
bb	1344	CAGGGTGACAGTGGTGGCCCCCTGATGTACCAATCTGACCACTGGCATGTGGTGGCATC	1403
yy	402	ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer	421
b	1404	GTTAGCTGGGGCTATGGGTGGGGGGCCCGAGCACCCAGGAGTATACCAAGGTCTCA	1463
yy	422	AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	435
b	1464	GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG	1505
RESULT 14			
X697206			
OCUS			
EFINITION			
Sequence 274 from Patent WO0078961.			
AX697206			
VERSION			
AX697206.1 GI:29498146			
KEYWORDS			
SOURCE			
Homo sapiens (human)			
ORGANISM			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
1			
AUTHORS			
Ferrara, N., Stewart, T.A., Williams, P.M., Baker, K.P., Desnoyers, L.,			
Eaton, D.L., Gao, W.Q., Pan, J., Botstein, D., Fong, S., Goddard, A.,			
Godowski, P.J., Gurney, A.L., Smith, V., Tumas, D., Wood, W.I.,			
Grimaldi, C.J., Hillan, K.J., Paoni, N.F., Roy, M.A. and Watanabe, C.K.			
Secreted and transmembrane polypeptides and nucleic acids encoding			
the same			
TITLE			
JOURNAL			
Patent: WO 0078961-A 274 28-DEC-2000;			
Genentech Inc. (US)			
FEATURES			
source			
1. .2063			
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
ASE COUNT			
477 a 591 c 576 g 419 t			
RIGIN			
Alignment Scores:			
red. No.: 3.17e-210 Length: 2063			
core: 2297.50 Matches: 429			
ercent Similarity: 98.85% Conservative: 0			
est Local Similarity: 98.85% Mismatches: 0			
uery Match: 98.10% Indels: 5			
B: 6 Gaps: 1			
US-09-607-745-2 (1-435) x AX697206 (1-2063)			
2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg			
yy	21		

Db	219	GATCCTGACAGTGATCAACCTCTGAACAGCCCTCGATGTCAAACCCCTGCGCAAAACCCCGT	278
Qy	22	IleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSerLeu	41
Db	279	ATCCCCATGGAGACCTTCAGAAAGGTGGGATCCCCATCATATAGCACTACTAGCCTG	338
Qy	42	AlaSerIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu	61
Db	339	CGGAGTATCATATTGTTGTCTCATCAAGGTGATTCTGGATAAATACTACTTCTCTC	398
Qy	62	CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys	81
Db	399	TGCGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGACTGT	458
Qy	82	ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla	101
Db	459	CCCTTGGGGGAGGACGAGGAGCACTGTCTCAAGAGCTTCCCGAAGGGCCTGCAGTGGCA	518
Qy	102	ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp	121
Db	519	GTCCGCCTCTCCAAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGAACTGG	578
Qy	122	PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet	141
Db	579	TTCTCTGCCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAGATG	638
Qy	142	GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp	161
Db	639	GGCTACAGC-----AGAGCTGTGGAGATTGGCCACAGCAGGATCTGGAT	683
Qy	162	ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys	181
Db	684	GTTGTTGAATCACAGAAACACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCCCTGT	743
Qy	182	LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro	201
Db	744	CTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGCCTGTGGGAAGAGCCTGAAGACCCC	803
Qy	202	ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln	221
Db	804	CGTGTGTTGGTGGGAGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTCAGCATCCAG	863
Qy	222	TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla	241
Db	864	TACGACAAACAGCACGCTCTGTGGAGGAGCATCTCTGGACCCCACTGGGTCTCTCACGGCA	923
Qy	242	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	261
Db	924	GCCCACTGCTTCAGGAACATACCGATGTGTTCACTGGAAGTGGGGCAGGCTCAGAC	983
Qy	262	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro	281
Db	984	AAACTGGGCAGTTCCCATCCCTGGCTGTGGCCAAAGATCATCATATTGAATTCAACCCC	1043
Qy	282	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	301
Db	1044	ATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTCCCACCTCACTTTCTCA	1103
Qy	302	GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	321
Db	1104	GGCAGATCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA	1163
Qy	322	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	341
Db	1164	CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACTG	1223
Qy	342	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	361
Db	1224	CTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGTCAATGCAGACGATCGGTACCG	1283
Qy	362	GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	381
Db	1284	GGGGAAGTCACCGAGAAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACCTGC	1343

Y 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
b 1344 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGGCATC 1403
Y 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
b 1404 GTTAGCTGGGGCTATGGCTCGGGGGCCCGAGCACCCAGGAGATATACACCAAGGTCTCA 1463
Y 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
b 1464 GCCTATCTCAACTGGATCTACATGTCTGGAAGGCTGAGCTG 1505

RESULT 15
C012752
OCUS BC012752 1793 bp mRNA linear PRI 20-AUG-2001
DEFINITION Homo sapiens, Similar to transmembrane protease, serine 4, clone
IMAGE:3623466, mRNA, partial cds.

CCESION BC012752 GI:15215322
ERSION
EYWORDS
OURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1793)
Strausberg,R.
Direct Submission
Submitted (15-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 24 Row: m Column: 19.

FEATURES
source Location/Qualifiers
1..1793
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3623466"
/tissue_type="Pancreas, adenocarcinoma"
/clone_lib="NIH_MGC_39"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
<1..1219
/codon_start=2
/product="Similar to transmembrane protease, serine 4"
/protein_id="AAH12752.1"
/db_xref="GI:15215323"

CDS
translation="ARGIIALLSLASIIIVVLIKVIIDKYFLCQQLHPIPRKQL
CDGELDCPLGEDEEHCKVSPFEGPAVAVRLSKDRSTQLVDSATGNWFSACDFNTEA
LAETACROMGYSSKPTFRAVEIGPDQDLVVEITENSQELMRNNSGPGCLSGILVSLH
CLACGSLKTPRVGGEASVDSWFWQVSIQYDKQHVCGSILDPHWLVLTAAHCFRKH
TDVENWKVRAGSKDLGSPSLAVAKIIIEFNPMYPKNDIALMKLOFPLTPSGTVRP
ICKPFFDEELTPATPLIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEV
TEKLMCAGIPEGVDTCQDSSGGLMYQSDQHWVVGIVSWGVCGGPSTPGVYTKVSA
YLNWIYNVWKAEL"

ASE COUNT 436 a 501 c 486 g 370 t
RIGIN

Alignment Scores:
Pred. No.: 8.49e-138 Length: 1793
Score: 2167.00 Matches: 402
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 92.53% Indels: 0
DB: 9 Gaps: 0
US-09-607-745-2 (1-435) x BC012752 (1-1793)
QY 34 IleIleIleAlaLeuLeuSerLeuAlaSerIleIleIleValValValLeuIleLysVal 53
Db 11 ATCATCATAGCACTACTAGCCCTGGCGAGTATCATCAITGTGGTTGTCTCATCAAGGTG 70
QY 54 IleLeuAspLysTyrTyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGln 73
Db 71 ATTCTGGATAAATACTACTTCTCTGGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAG 130
QY 74 LeuCysAspGlyGluLeuAspCysProLeuGlyGluAspGluHisCysValLysSer 93
Db 131 CTGTGTGACGGAGAGCTGGACTGTCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGC 190
QY 94 PheProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnVal 113
Db 191 TTCCCGAAGGGCCTGCAGTGGCAGTCCGCTCTCCAGGACCGATCCACACTGCAGGTG 250
QY 114 LeuAspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeu 133
Db 251 CTGGACTCGGCCACAGGGAACTGTCTCTGCTGTTTTCGACAACTTCACAGAAGCTCTC 310
QY 134 AlaGluThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGlu 153
Db 311 GCTGAGACAGCCTGTAGGCAGATGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAG 370
QY 154 IleGlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArg 173
Db 371 ATTGGCCCGACACAGGATCTGGATGTGTGAATCAGAGAAACAGCCAGGAGCTTCGC 430
QY 174 MetArgAsnSerSerGlyProCysLeuSerGlySerLeuValSerLeuHisCysLeuAla 193
Db 431 ATGCGGAACCTCAAGTGGGCCCTGTCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGCC 490
QY 194 CysGlyLysSerLeuLysThrProArgValValGlyGlyGluGluAlaSerValAspSer 213
Db 491 TGTGGGAAGAGCCTGAAGACCCCGTGTGTGGTGGGAGGAGGCTCTGTGTGATTCT 550
QY 214 TrpProTrpGlnValSerIleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeu 233
Db 551 TGGCCTTGGCAGGTCAGCATCCAGTACGACAAACAGCACGCTGTGTGGAGGAGCATCTCTG 610
QY 234 AspProHisTrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsn 253
Db 611 GACCCCACTGGGTCTCTACGGCAGCCCACTGTCTTACGAGAAACATACCGATGTGTTCAC 670
QY 254 TrpLysValArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLys 273
Db 671 TGGAAGGTGGGGCAGGCTCAGACAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCCAAG 730
QY 274 IleIleIleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLys 293
Db 731 ATCATCATATTGAATTCAACCCCATGTATACCCCAAGACAATGACATCGCCCTCATGAAG 790
QY 294 LeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAsp 313
Db 791 CTGCAGTTCACACTCACTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCITTGAT 850
QY 314 GluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsn 333
Db 851 GAGGAGCTCACTCCAGCCACCCCACTCTGGATCATTTGATGGGGCTTTACGAAGCAGAAT 910
QY 334 GlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArg 353
Db 911 GGAGGGAAGATGTCTGACATACTGTCTGCAGGCGCTCAGTCCAGGTCATTGACAGCACCGG 970

```
Y      354 CysAsnAlaAspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIle 373
b      |||||||
b      971 TGCAATGCAGACGATGCGTACCAGGGGAAGTCACCGAGAGATGATGTGTGCAGGCATC 1030
Y      374 ProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSer 393
b      |||||||
b      1031 CCGGAAGGGGTGTGGACACCTGCCAGGGTGACAGTGGTGGGCCCCCTGATGTACCAATCT 1090
Y      394 AspGlnTrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThr 413
b      |||||||
b      1091 GACCAGTGGCATGTGTGGGCATCGTTAGCTGGGGCTATGGCTGCGGGGCCCCGAGCAC 1150
Y      414 ProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAla 433
b      |||||||
b      1151 CCAGGAGTATACACCAAGGTCTCAGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCT 1210
Y      434 GluLeu 435
b      |||||
b      1211 GAGCTG 1216
```

earch completed: November 23, 2003, 10:39:57
ob time : 5540.45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

run on: November 23, 2003, 07:53:55 ; Search time 87.359 Seconds
(without alignments)
2197.846 Million cell updates/sec

title: US-09-607-745-2
perfect score: 2342
sequence: 1 MDPDSQPLNSLDVKPLRKP.....VYTKVSAYLNWIYNVWKAEI 435

scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

searched: 569978 seqs, 220691566 residues

total number of hits satisfying chosen parameters: 1139956

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

ommand line parameters:

MODEL=frame+ p2n.model -DEV=xlh
Q=/cgn2_1/USPTO_spool/US09607745/runat 21112003 144346 22244/app query.fasta_1.1038
DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosom62 -TRANS=human40.cdi
LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US09607745 @CGN 1 1 76 @runat 21112003 144346 22244 -NCPU=6 -ICPU=3
NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase : Issued Patents NA:

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Match	Length	DB ID	Description
1	2338	99.8	2038	3	US-09-008-271A-18
2	2324	99.2	2079	4	US-09-656-002-1
3	677.5	28.9	2479	3	US-09-342-749-29
4	677.5	28.9	2479	4	US-09-691-840-29
5	676.5	28.9	1479	3	US-09-342-749-1
6	676.5	28.9	1479	4	US-09-691-840-1
7	650.5	27.8	2413	3	US-09-518-046-1
8	627	26.8	2544	3	US-09-518-046-3
9	621.5	26.5	2416	3	US-09-261-416-1
10	584.5	25.0	1783	3	US-09-510-738A-188
11	584.5	25.0	1783	4	US-09-861-966-188
12	584.5	25.0	2363	4	US-09-742-703-3

13	580	24.8	1615	4	US-09-820-002-1	Sequence 1, Appli
14	578	24.7	1605	2	US-09-000-846-1	Sequence 1, Appli
15	576	24.6	1460	4	US-09-370-838-80	Sequence 80, Appl
16	576	24.6	1517	1	US-08-508-448C-15	Sequence 15, Appl
17	576	24.6	2790	4	US-09-370-838-79	Sequence 79, Appl
18	574	24.5	1462	4	US-09-370-838-55	Sequence 55, Appl
19	571	24.4	2581	1	US-08-200-900A-1	Sequence 1, Appli
20	571	24.4	2581	5	PCT-US94-00616-1	Sequence 1, Appli
21	558.5	23.8	1077	3	US-08-807-151-2	Sequence 2, Appli
22	558.5	23.8	1077	4	US-09-478-957-2	Sequence 2, Appli
23	531.5	22.7	696	1	US-08-508-448C-24	Sequence 24, Appl
24	516.5	22.1	901	1	US-08-508-448C-9	Sequence 9, Appli
25	503.5	21.5	959	4	US-09-023-942A-25	Sequence 25, Appl
26	496	21.2	1613	4	US-09-387-375-1	Sequence 1, Appli
27	483	20.6	1100	4	US-09-023-942A-5	Sequence 5, Appli
28	478	20.4	1081	3	US-09-008-271A-15	Sequence 15, Appl
29	477	20.4	1094	4	US-09-023-942A-3	Sequence 3, Appli
30	472.5	20.2	1430	4	US-09-386-629-1	Sequence 1, Appli
31	467.5	20.0	1110	4	US-09-386-653A-1	Sequence 1, Appli
32	466.5	19.9	1225	4	US-09-734-675-1	Sequence 1, Appli
33	463	19.8	3147	2	US-09-027-337-1	Sequence 1, Appli
34	463	19.8	3147	4	US-09-644-600-1	Sequence 1, Appli
35	463	19.8	3147	4	US-09-644-600-18	Sequence 18, Appl
36	462.5	19.7	980	4	US-09-023-942A-30	Sequence 30, Appl
37	462.5	19.7	1212	4	US-09-620-312D-431	Sequence 431, App
38	458.5	19.6	1130	4	US-09-387-375-8	Sequence 8, Appli
39	458	19.6	1739	2	US-08-681-151-2	Sequence 2, Appli
40	449.5	19.2	1165	4	US-09-023-942A-28	Sequence 28, Appl
41	445	19.0	933	4	US-09-023-942A-29	Sequence 29, Appl
42	443.5	18.9	1130	4	US-09-386-653A-8	Sequence 8, Appli
43	443	18.9	2296	1	US-07-750-080A-18	Sequence 18, Appl
44	443	18.9	2296	3	US-08-651-472-18	Sequence 18, Appl
45	443	18.9	2296	3	US-08-358-928-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-09-008-271A-18
; Sequence 18, Application US/09008271A
; Patent No. 6203979
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: COLNNOT13
CLONE: 1337018

SEQUENCE DESCRIPTION: SEQ ID NO: 18:

S-09-008-271A-18

Alignment Scores:

red. No.: 6.49e-244 Length: 2038
core: 2338.00 Matches: 434
percent Similarity: 100.00% Conservativity: 1
est Local Similarity: 99.77% Mismatches: 0
Query Match: 99.83% Indels: 0
GB: 3 Gaps: 0

JS-09-607-745-2 (1-435) x US-09-008-271A-18 (1-2038)

Qy	1	MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro	20
Db	200	ATGGATCCTGACAGTGATCAACCTCTGAACAGCCCTCGATGTCAACCCCTGCGCAAAACC	259
Y	21	ArgIleProMetGluThrPheArgLysValGlyLeuProIleIleLeuLeuSer	40
b	260	CGTATCCCCATGGAGACCTTCAGAAAAGGTGGGATCCCCCATCATCATAGCACTACTGAGC	319
Y	41	LeuAlaSerIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPhe	60
b	320	CTGGCGAGTATCATCTTGTGTCTCCTCATCAAGGTGATTCGTGATAAATACTACTTC	379
Y	61	LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp	80
Db	380	CTCTGGCGGCAGCCCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGAGAGCTGGAC	439
Y	81	CysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaVal	100
b	440	TGTCCCTTGGGGAGGAGCAGGAGGACACTGTGTCAAGAGCTTCCCCGAAGGCCCTGCAGTG	499
Y	101	AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn	120
b	500	GCAGTCCGCCTCTCCAAAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGGAAC	559
Y	121	TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln	140
b	560	TGTTCTCTGCCTGTTTCGACAACTTCACAGAGCTCTCGTGTGAGACGCTGTAGGCAG	619
Y	141	MetGlyTyrSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu	160
b	620	ATGGGTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCATCTG	679
Y	161	AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro	180
b	680	GATGTTGTTGAATCACAGAAACAGCCAGGAGTTTCGATTCGGGAATCAAGTGGGCC	739
Y	181	CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr	200
b	740	TGTTCTCTCAGGCTCCCTGTCTCCCTGCACCTGTTCCTGTGGGAGAGCCCTGAAGACC	799
Y	201	ProArgValValGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIle	220
b	800	CCCCGTGTGTGGTGGGAGGAGGCCCTCTGTGGATTCTTGGCCTTGGCAGGTCTGATC	859
Y	221	GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr	240
b	860	CAGTACGACAAACAGCACCTCTGTGGAGGGAGCATCTTGACCCCCCACTGGTCTCTCAG	919

Qy	241	AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer	260
Db	920	GCAGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGTGGGGCAGGCTCA	979
Qy	261	AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsn	280
Db	980	GACAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCAAGATCATCATCATGAATTCAC	1039
Qy	281	ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe	300
Db	1040	CCCATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGAGTCCCACCTCTTC	1099
Qy	301	SerGlyThrValArgProIleCysLeuProphePheAspGluGluLeuThrProAlaThr	320
Db	1100	TCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC	1159
Qy	321	ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle	340
Db	1160	CCACTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATA	1219
Qy	341	LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyr	360
Db	1220	CTGCTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGCAATGCAGACGATGCGTAC	1279
Qy	361	GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr	380
Db	1280	CAGGGGAAGTCACCGAGAAGATGATGTGTGAGGCATCCCCGAAGGGGTGTGGACACC	1339
Qy	381	CysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGly	400
Db	1340	TGCCAGGTGACAGTGTGGGCCCTCTGATGTACCAATCTGACCCAGTGGCATGTGTGGGC	1399
Qy	401	IleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVal	420
Db	1400	ATCGTTAGCTGGGCTATGGTGTGGGGCCCTGATGTACCAATCTGACCCAGGAGTATACACCAAGGTC	1459
Qy	421	SerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	435
Db	1460	TCAGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG	1504

RESULT 2

US-09-656-002-1

; Sequence 1, Application US/09656002

; Patent No. 6455668

; GENERAL INFORMATION:

; APPLICANT: Mack, David

; APPLICANT: Gish, Kurt

; APPLICANT: Wilson, Keith

; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND

; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS

; FILE REFERENCE: A-69108/DJB/JJD/AMS

; CURRENT APPLICATION NUMBER: US/09/656,002

; CURRENT FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: US 09/525,993

; PRIOR FILING DATE: 2000-03-15

; PRIOR APPLICATION NUMBER: US 09/493,444

; PRIOR FILING DATE: 2000-01-28

; PRIOR APPLICATION NUMBER: PCT/US 00/07044

; PRIOR FILING DATE: 2000-03-15

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1

; LENGTH: 2079

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-656-002-1

Alignment Scores:

Pred. No.: 2.22e-242 Length: 2079
Score: 2324.00 Matches: 434
Percent Similarity: 99.77% Conservativity: 0
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.23% Indels: 1

B: 4 Gaps: 0

S-09-607-745-2 (1-435) x US-09-656-002-1 (1-2079)

2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
217 GATCCTGACAGTGATCAACCTCTGAACAGCCCTCGATGTCAAACCCCTGCGCAACCCCGT 276

22 IleProMetGluThrPheArgLys-ValGlyIleProIleIleIleAlaLeuLeuSerLe 41
277 ATCCCATGGAGACCTTCAGAAAGTGTGGGATCCCCATCATCATAGCACTACTAGCCT 336

41 uAlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLe 61
337 GCGGAGTATCATATTGGTGTCTCATCAAGGTGATTCGTGATAAATACTACTTCTCT 396

61 uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCy 81
397 CTGCGGCAGCCTCTCCACTTCATCCCAGGAAGCAGCTGTGTGACGAGAGCTGGACTG 456

81 sProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAl 101
457 TCCCTTGGGGAGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAGGCCCTGCAGTGGC 516

101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
517 AGTCCGCCTCTCCAAGGACCGATCCACACTGCGAGTGTGGACTCGGCCACAGGGAAC 576

121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
577 GTTCTCTGCCTGTTTCGACAACCTTCACAGAGCTCTCGCTGAGACGCCCTGTAGGCAG 636

141 tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs 161
637 GGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCCCAGACAGGATCT 696

161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181
697 TGTGTGTTGAAATCAGAAAAACAGCCAGGAGTTCGCATGCGGAACCTCAAGTGGCCCTG 756

181 sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr 201
757 TCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTTCCTGTGGGAGAGCCTGAAGACCCC 816

201 oArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIleGl 221
817 CCGTGTGTGGTGGGAGGAGGCCCTCTGTGATTCCTTGGCCTTGGCAGGTGAGCATCCA 876

221 nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl 241
877 GTACGACAAACAGCAGCTGTGTGGAGGAGCATCTGGACCCCACTGGGTCTCTCACGGC 936

241 aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261
937 AGCCCACTGCTTCAGAAACATACCGATGTGTCAACTGGAAGGTGCGGGCAGGCTCAGA 996

261 pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPr 281
997 CAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAGATCATCATCATTTGAATTCAACC 1056

281 oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301
1057 CATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGAGTTCCTCCACTCACTTCTC 1116

301 rGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPr 321
1117 AGGCACAGTCAGGCCCATCTGTCTGCCCCTCTTTGATGAGGAGCTCACTCCAGCCACCC 1176

321 oLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLe 341
1177 ACTCTGGATCATTTGGATGGGCTTACGAACAGAAATGGAGGGAGATGTCTGACATACT 1236

341 uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGl 361

Db 1237 GCTGCAGGCGTCAGTCCAGTCAATGACAGCACACGGTGCAATGACAGCATGCGTACCA 1296

QY 361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381
1297 GGGGAAGTCACCGAAGATGATGTGTGAGGCATCCCCGGAAGGGGTGTGGACACCTG 1356

QY 381 sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIl 401
1357 CCAGGTGACAGTGGTGGCCCCCTGATGTACCAATCTGACCAAGTGGCATGTGTGGGCAT 1416

QY 401 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValse 421
1417 CGTTAGCTGGGCTATGGCTGCGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTC 1476

QY 421 rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
1477 AGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1519

RESULT 3

US-09-342-749-29
; Sequence 29, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-342-749-29

Alignment Scores:
Pred. No.: 1.81e-63 Length: 2479
Score: 677.50 Matches: 151
Percent Similarity: 54.05% Conservative: 56
Best Local Similarity: 39.43% Mismatches: 127
Query Match: 28.93% Indels: 49
DB: 3 Gaps: 13

US-09-607-745-2 (1-435) x US-09-342-749-29 (1-2479)

QY 75 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPhe 94
453 TGTGATGGCGTGTACACTGCCCGCGGGGAGGACGAGAAATCGGTGTGTTCGCCTCTAC 512

QY 95 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 114
513 -----GGACCA-----AACTTCATCCTTCAGATGTAC 539

QY 115 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla 134
540 TCATCTCAGAGGAAGTCTCTGGCACCTGTGTGCCAAGACGACTGGAACGAGAACTACGGG 599

QY 135 GluThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIle 154
600 CGGGCGGCTGCAGGGACATGGGCTAT-----AAGATAATTTTACTCTAGCCAA--- 650

QY 155 GlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMet 174
651 -----GGATAGTGGATGACAGCGGATCCACCAGCTTTATGAAACTG 692

QY 175 ArgAsnSerSerGly-----ProCysLeu 182

```

b 693 AACACAAAGTCCGGCAATGTCGATATCTATAAAAACTGTACCACAGTGCCTGTTCT 752
y 183 SerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeu-----Lys 199
b 753 TCAAAAGCAGTGGTTCTTTACGCTGTTTAGCCTGCGGGTCAACTTGAACTCAAGCCGC 812
y 200 ThrProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSer 219
b 813 CAGAGCAGGATCGTGGCGGTGAGAGCGCGCTCCCGGGGCTGGCCCTGGCAGGTGAGC 872
y 220 IleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu 239
b 873 CTGCACCTCCAGAACGTCACGCTGCGGAGGCTCCATCATCACCCCGAGTGCATCGTG 932
y 240 ThrAlaAlaHisCysPheArgLysHis---ThrAspValPheAsnTrpLysValAla 258
b 933 ACAGCCGCCCACTGCGTGGAAAAACCTCTTAACAATCCATGGCATTTGGACGCATTGCG 992
y 259 GlySerAspLysLeuGlySerPhe-----ProSerLeuAlaValAlaLysIle 274
b 993 GGGATTTTGAGA---CAATCTTTCATGTTCTATGGAGCCGGATACCAAGTACAAAAGTG 1049
y 275 IleIleIleGluPheAsnProMetTyr-----ProLysAspAsnAspIleAlaLeu 291
b 1050 ATT-----TCTCATCCAAATATGACTCCAAGACCAAGAACRAATGACATTGCGCTG 1100
y 292 MetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhe 311
b 1101 ATGAAGCTGCAGAACCTCTGACTTTCAACGACCTAGTGAACCAAGTGTGTCTGCCCAAC 1160
y 312 PheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLys 331
b 1161 CCAGGCATGATGCTGCAGCCAGAACAGCTCTGCTGGATTCCGGTGGGGGCCACCGAG 1220
y 332 GlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSer 351
b 1221 GAGAAA---GGGAAGACCTCAGAACTGCTGAACGCTGCCAAGGTGCTTCTCATTTGAGACA 1277
y 352 ThrArgCysAsnAlaAspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAla 371
b 1278 CAGAGATGCAACAGCAGATATGTCTATGACAACTGATCACCAGCCATGATCTGTGCC 1337
y 372 GlyIleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyr 391
b 1338 GGCTTCCTGCAGGGGAACGTCGATTCTTGCCAGGTGACAGTGGAGGCTCTGCTCACT 1397
y 392 GlnSerAspGln---TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGly 410
b 1398 TCGAACACAATAATCTGTTGGCTGATAGGGGATACAAAGCTGGGGTCTGGCTGTGCCAAA 1457
y 411 ProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnVal 430
b 1458 GCTTACAGACAGGAGGTGTACGGGAATGTGATGTTATCACGGAATGATTTATCGACAA 1517
y 431 TrpLysAla 433
b 1518 ATGAAGGCA 1526
```

RESULT 4

```

US-09-691-840-29
; Sequence 29, Application US/09691840
; Patent No. 6444419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRPSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/691,840
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
```

```

; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-691-840-29

Alignment Scores:
Pred. No.: 1,81e-63 Length: 2479
Score: 677.50 Matches: 151
Percent Similarity: 54.05% Conservative: 56
Best Local Similarity: 39.43% Mismatches: 127
Query Match: 28.93% Indels: 49
DB: 4 Gaps: 13

US-09-607-745-2 (1-435) x US-09-691-840-29 (1-2479)

Qy 75 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluHisCysValLysSerPhe 94
Db 453 TGTGATGGCGTGTACACTGCCCGGGGGAGGACGAGAAATCGGTGTGTTCGCCTCTAC 512
Qy 95 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 114
Db 513 -----GGACCA-----AATTCTATCTTCAGATGTAC 539
Qy 115 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla 134
Db 540 TCATCTCAGAGGAAGTCTGGCACCCCTGTGTGCCAAGACGACTGGAAACGAGAACTACGGG 599
Qy 135 GluThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIle 154
Db 600 CGGCGCGCCTGCAGGACATGGGCTAT-----AAGAATAATTTTACTCTAGCCAA--- 650
Qy 155 GlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMet 174
Db 651 -----GGATAGTGGATGACAGCGGATCCACCAGCTTTATGAAACTG 692
Qy 175 ArgAsnSerSerGly-----ProCysLeu 182
Db 693 AACACAAAGTCCGGCAATGTCGATATCTATAAAAACTGTACCACAGTGCCTGTTCT 752
Qy 183 SerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeu-----Lys 199
Db 753 TCAAAAGCAGTGGTTCTTTACGCTGTTTAGCCTGCGGGGTCAACTTGAACTCAAGCCGC 812
Qy 200 ThrProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSer 219
Db 813 CAGAGCAGGATCGTGGCGGTGAGAGCGCGCTCCCGGGGCTGGCCCTGGCAGGTGAGC 872
Qy 220 IleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu 239
Db 873 CTGCACCTCCAGAACGTCACGCTGCGGAGGCTCCATCATCACCCCGAGTGCATCGTG 932
Qy 240 ThrAlaAlaHisCysPheArgLysHis---ThrAspValPheAsnTrpLysValAla 258
Db 933 ACAGCCGCCCACTGCGTGGAAAAACCTCTTAACAATCCATGGCATTTGGACGCATTGCG 992
Qy 259 GlySerAspLysLeuGlySerPhe-----ProSerLeuAlaValAlaLysIle 274
Db 993 GGGATTTTGAGA---CAATCTTTCATGTTCTATGGAGCCGGATACCAAGTACAAAAGTG 1049
Qy 275 IleIleIleGluPheAsnProMetTyr-----ProLysAspAsnAspIleAlaLeu 291
Db 1050 ATT-----TCTCATCCAAATATGACTCCAAGACCAAGAACRAATGACATTGCGCTG 1100
Qy 292 MetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhe 311
Db 1101 ATGAAGCTGCAGAACCTCTGACTTTCAACGACCTAGTGAACCAAGTGTGTCTGCCCAAC 1160
```


817 CTGCCAGCTCCAGAACGTCCACAGTGTGCGGAGGTCCCATCATCACCCCGAGTGGATCGTG 876

240 ThrAlaAlaHisCysPheArgLysHis---ThrAspValPheAsnTrpLysValArgAla 258
|||||
877 ACAGCCGCCCACTGCTGGAAAAACCTCTTAACAATCCATGGCATTTGGACGGCATTTGCG 936
|||||

259 GlySerAspLysLeuGlySerPhe-----ProSerLeuAlaValAlaLysIle 274
|||
937 GGGATTTTGAGA--CAATCTTTTCATGTTCTATGGAGCCGGATACCAAGTAGAAAAAGTG 993
|||

275 IleIleIleGluPheAsnProMetTyr-----ProLysAspAsnAspIleAlaLeu 291
|||
994 ATT-----TCTCATCCAAATTATGACTCCAGACCAAGAACATGACATTGCGCTG 1044
|||

292 MetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhe 311
|||||
1045 ATGAAGCTGCAGAGCCCTCTGACTTTTCAACGACCTAGTGAACCCAGTGTGTCTGCCCAAC 1104
|||||

312 PheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLys 331
|||
1105 CCAGGCATGATGCTGCAGCCAGAACACAGCTCTGCTGGATTTCGGGTGGGGGCCACCGAG 1164
|||

332 GlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSer 351
:::
1165 GAGAAA---GGGAAGACCTCAGAAGTGTGAACGCTGCCAAGGTGCTTCTCATTTGAGACA 1221
|||

352 ThrArgCysAsnAlaAspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAla 371
|||
1222 CAGAGATGCAACAGCAGAGATATGCTATGACAACTGTATCACACCGCCATGATCTGTGCC 1281
|||

372 GlyIleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyr 391
|||
1282 GGCTTCCCTGCAGGGGAACGTGCGATTCTTGCAGGGTGACAGTGGAGGGCCCTCTGGTCACT 1341
|||

392 GlnSerAspGln---TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGly 410
:::
1342 TCGAAGAACAATATCTGGTGGCTGATAGGGGATACAAGCTGGGTCTTCTGGCTGTGCCAAA 1401
|||

411 ProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnVal 430
|||
1402 GCTTACAGACCAGAGTGTACGGGAATGTATGGTATTTCACGGACTGGATTATCGACAA 1461
|||

431 TrpLysAlaGlu 434
:::
1462 ATGAGGGCAGAC 1473

RESULT 6

6-09-691-840-1
Sequence 1, Application US/09691840

FACEID NO. 044419
GENERAL INFORMATION.

GENERAL INFORMATION:
 APPLICANT: Wong, Alexander K.C.
 APPLICANT: Tavtigian, Sean V.
 APPLICANT: Teng, David H.-F.
 APPLICANT: Myriad Genetics, Inc.
 TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
 FILE REFERENCE: 2318-202
 CURRENT APPLICATION NUMBER: US/09/691,840
 CURRENT FILING DATE: 2000-10-18
 PRIOR APPLICATION NUMBER: US/09/342,749
 PRIOR FILING DATE: 1999-06-29
 PRIOR APPLICATION NUMBER: US 60/091,044
 PRIOR FILING DATE: 1998-06-29
 NUMBER OF SEQ ID NOS: 33

```

SEQ ID NO 1
LENGTH: 1479
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1476)
NAME/KEY: conflict

```

```

; LOCATION: (724)
; OTHER INFORMATION: Listed as T in GenBank Accession NO. U75329
; NAME/KEY: conflict
; LOCATION: (985)
; OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
; NAME/KEY: conflict
; LOCATION: (1347)
; OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
; NAME/KEY: conflict
; LOCATION: (1466)
; OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329
; NAME/KEY: conflict
; LOCATION: (1471)
; OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329.
; NAME/KEY: allele
; LOCATION: (478)
; OTHER INFORMATION: This base can be G or A with G being the more
; OTHER INFORMATION: common allele. The codon will change from Val to
; OTHER INFORMATION: Met.
; NAME/KEY: allele
; LOCATION: (777)
; OTHER INFORMATION: This base can be C or T with C being the more
; OTHER INFORMATION: common allele. The codon is unaffected with both
; OTHER INFORMATION: alleles encoding Gly.
; NAME/KEY: allele
; LOCATION: (768)
; OTHER INFORMATION: This base can be C or T with C being the more
; OTHER INFORMATION: common allele. This is a silent polymorphism.
; NAME/KEY: allele
; LOCATION: (834)
; OTHER INFORMATION: This base can be C or T with C being the more
; OTHER INFORMATION: common allele. This is a silent polymorphism.
; NAME/KEY: allele
; LOCATION: (625)
; OTHER INFORMATION: This base can be T or A with T being the more
; OTHER INFORMATION: common allele. The codon will change from Phe to Ile
US-09-691-840-1

Alignment Scores:
Pred. No.: 1.04e-63 Length: 1479
Score: 676.50 Matches: 150
Percent Similarity: 53.91% Conservative: 57
Best Local Similarity: 39.06% Mismatches: 128
Query Match: 28.89% Indels: 49
DB: 4 Gaps: 13

```

US-09-607-745-2	(1-435)	x	US-09-691-840-1	(1-1479)	
QY	75	CysAspGlyGluLeuAspCysProLeuGlyGluAspGluHisCysValLysserPhe	94		
Db	397	TGTGATGGCGTGTCACTGCCCGCGGGAGACGAGAATCGGTGTGTTCCGCCTCTAC	456		
QY	95	ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu	114		
Db	457	-----GGACCA-----AATTTCATCCTTCAGGTGTAC	483		
QY	115	AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla	134		
Db	484	TCATCTCAGAGGAAGTCTCGCACCTGTGTGCCAGACGACTGGAACGAGAACTACGGG	543		
QY	135	GluThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIle	154		
Db	544	CGGCGGCCTGCAGGGACATGGGCTAT-----AAGAATAATTTTTACTCTAGCCAA---	594		
QY	155	GlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMet	174		
Db	595	-----GGAATAGTGGATGACAGCGGATCCACCAGCTTTATGAAACTG	636		
QY	175	ArgAsnSerSerGly-----proCysLeu	182		
Db	637	AACACAAGTGCCGGCAATGTCGATATCTATAAAAACTGTACCACAGTGATGCTGTTCT	696		
QY	183	SerGlySerLeuValserLeuHisCysLeuAlaCysGlyLvsserLeu-----lvs	199		

1020 -----TACCACAGCAAGTACAAGCCAAAGAGGCTGGGCAATGACATCGCCCTTATGAAG 1073
294 LeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAsp 313
1074 CTGGCCGGGGCCACTCAGTTCAATGAATGATCCAGCCTGTGTGCTGCTGCCCACTTGAA 1133
314 GluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsn 333
1134 GAGAACTTCCCGATGGAAGAGTGTCTGGACGTGAGGATGGGGGCCACA---GAGGAT 1190
334 GlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArg 353
1191 GGAGGTGACGCTCCCTGTCTGAACACCGCGGCGCTCCCTTTGATTCCAACAAGATC 1250
354 CysAsnAlaAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIle 373
1251 TGCAACCACAGGACGTGTACGGTGGCATCATCTCCCTCCATGCTCTGCGGGGCTAC 1310
374 ProGluGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSer 393
1311 CTGACGGGTGGCGTGAACAGCTGCCAGGGGACAGCGGGGGCCCTGGTGTGTCAAGAG 1370
394 AspGln---TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSer 412
1371 AGGAGGCTGTGGAAGTGTAGTGGAGCGACCAAGCTTTGGCATCGGCTGGCAGAGGTGAAC 1430
413 ThrProGlyValThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLys 432
1431 AAGCCTGGGTGTACACCCGTGTACCTCCTTCTGGACTGGATCCACGAGCAGATGGAG 1490
433 AlaGluLeu 435
1491 AGAGACCTA 1499
RESULT 8
S-09-518-046-3
Sequence 3, Application US/09518046
Patent No. 6294663
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
APPLICANT: Underwood, Lowell J.
TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
FILE REFERENCE: D6192CIP
CURRENT APPLICATION NUMBER: US/09/518,046
CURRENT FILING DATE: 2000-03-02
EARLIER APPLICATION NUMBER: 09/261,416
EARLIER FILING DATE: 1999-03-03
NUMBER OF SEQ ID NOS: 153
SEQ ID NO 3
LENGTH: 2544
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
OTHER INFORMATION: entire cDNA sequence of TADG-12 variant gene
S-09-518-046-3
Lignment Scores:
red. No.: 5.75e-58 Length: 2544
core: 627.00 Matches: 160
ercent Similarity: 48.14% Conservative: 73
est Local Similarity: 33.06% Mismatches: 150
uey Match: 26.77% Indels: 102
B: 3 Gaps: 18
S-09-607-745-2 (1-435) x US-09-518-046-3 (1-2544)
y 22 IleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSerLeu 41
b 273 CTGCCATTGAAGTTTTTCCCAATCATCGTATGGGGATCATTCGATTG----- 323

QY 42 AlaSerIleIleValValValLeuIleLysValIleLeuAsp-----LysTyr 58
Db 324 -----ATATTAGCACTGGCATTTGGTCTGGGCAATCCACTTCGACTGCTCAGGAAGTAC 377
QY 59 -----TyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeu 74
Db 378 AGATGTCGCTCATCCTTTAAGTGT-----ATCGAGCTGATAACTCGA----- 419
QY 75 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluHisCysValLysSerPhe 94
Db 420 TGTGACGGAGTCTCGGATTGCAAAAGACGGGGAGGACGAGTACCGCTGT----- 467
QY 95 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 114
Db 468 -----GTCCGGTGGGTGGTGTGTCAGAAATGCCGTGCTCCAGGTGTTTC 506
QY 115 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla 134
Db 507 ACAGCTGCT-----TCGTGGAAGACCATGTGCTCCGATGACTGGAAGGGTCACTACGCA 560
QY 135 GluThrAlaCysArgGlnMetGlyTyrSerSer----- 145
Db 561 AATGTTGCTGTGCCCACTGGGTTTCCCAAGCTATGTAAAGTTCAGATAAACCTCAGAGTG 620
QY 146 -----LysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAspVal 162
Db 621 AGCTCGCTGGAGGGGCGAGTTCCGGGAGGAGTTTGTGTCCATCGATCACCTCTTGCCAGAT 680
QY 163 ValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCysLeu 182
Db 681 GACAAGGTGACTGCATTACACCCTCAGTATATGTGAGGAGGGA-----TGTGCC 731
QY 183 SerGlySerLeuValSerLeuHisCysLeuAlaCysGly---LysSerLeuLysThrPro 201
Db 732 TCTGGCCACGTGGTTACCTTGAGTGCACAGCCTGTGTCATAGAGGGGTACAGTCA 791
QY 202 ArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
Db 792 CGCATCGTGGTGGAAACATGTCTTGTCTCTCGCAGTGGCCCTGGCAGGGCCGCTCAG 851
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
Db 852 TTCCAGGGTACCACCTGTGCGGGGGCTCTGTCTATCAGCCCCCTGTGGATCATCACTGCT 911
QY 242 AlaHis----- 243
Db 912 GCACACTGTGTTTATGAGATTGTAGCTCCTAGAGAAAGGGCAGACAGAGAGGAAGAAAG 971
QY 244 -----CysPheArgLysHisThrAspValPhe----- 252
Db 972 CTCCTGTGTGGAGGAAACCCACAAAATGAA-AGGACCTAGACCTTCCCATAGCTAATT 1030
QY 253 -----AsnTrpLysValArg 257
Db 1031 CCAGTGGACCATGTTATGGCAGATACAGGCTTGTACCTCCCCAAGTCATGGACCATCCAG 1090
QY 258 AlaGlySerAspLysLeu-----GlySerPheProSerLeuAlaValAlaLysIleIle 275
Db 1091 GTGGGTCTAGTTTCCCTGTTGGACAATCCAGCCCCATCCACTTGGTGGAGAGATTGTC 1150
QY 276 IleIleGluPheAsnProMetTyr---ProLys-----AspAsnAspIleAlaLeuMet 292
Db 1151 -----TACCACAGCAAGTACAAAGCCAAAGAGGCTGGGCAATGACATCGCCCTATG 1201
QY 293 LysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePhe 312
Db 1202 AAGCTGGCCGGGCCACTCACGTTCAATGAATGATCCAGCCTGTGTGCTGCCCAACTCT 1261
QY 313 AspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGln 332
Db 1262 GAAGAGAACTTCCCCGATGGAAAAGTGTCTGGAGCTGAGGATGGGGGGCCACA---GAG 1318
QY 333 AsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThr 352

1319 GATGGAGGTGACGCTCCCTGTCTGAACACGCGGCGTCCCTTGATTCCAAACAG 1378
353 ArgCysAsnAlaAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGly 372
1379 ATCTGCAACACACAGGACGTGTACGGTGGCATCATCTCCCTCCATGCTCTGCGCGGC 1438
373 IleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGln 392
1439 TACCTGACGGGTGGCTGGACAGCTGCCAGGGGACAGCGGGGGCCCTGTTGTCTCAA 1498
393 SerAspGln---TyrHisValValGlyIleValSerTyrGlyTyrGlyCysGlyGlyPro 411
1499 GAGAGGAGCTGTGGAAGTTAGTGGAGCGCAGCAGCTTTGGCATCGGCTCGCAGAGGTG 1558
412 SerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTyrPheTyrAsnValTrp 431
1559 AACAAAGCCTGGGTGTACACCCGTGTACCTCTCTCGACTGGATCCACGAGCAGATG 1618
432 LysAlaGluLeu 435
1619 GAGAGAGACCTA 1630

RESULT 9
S-09-261-416-1
Sequence 1, Application US/09261416A
Patent No. 6291663
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
APPLICANT: Underwood, Lowell J.
TITLE OF INVENTION: TADG-12: A No. 6291663el Transmembrane Serine Protease
TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
FILE REFERENCE: D6192
CURRENT APPLICATION NUMBER: US/09/261,416A
CURRENT FILING DATE: 1999-03-03
NUMBER OF SEQ ID NOS: 14
SEQ ID NO 1
LENGTH: 2416
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
LOCATION: 144..1511
OTHER INFORMATION: CDS
S-09-261-416-1

Alignment Scores:
red. No.: 2,1e-57 Length: 2416
core: 621,50 Matches: 158
percent Similarity: 51.57% Conservative: 72
est Local Similarity: 35.43% Mismatches: 150
query Match: 26.54% Indels: 66
B: 3 Gaps: 18

S-09-607-745-2 (1-435) x US-09-261-416-1 (1-2416)
22 IleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuSerLeu 41
273 CTGCCATTGAAGTTTTCCTCAATCATCGTTCATTGGGGATCATGTCATTG----- 323
42 AlaSerIleIleValValValLeuLysValIleLeuAsp-----LysTyr 58
324 -----ATATTAGACTGGCCATTGGTCTGGGCATCCACTTCGACTGCTCAGGGAAGTAC 377
59 -----TyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeu 74
378 AGATGTCGCTCATCTTTAAGTGT-----ATCAGCTGATAACTCGA----- 419
75 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPhe 94
420 TGTGACGGAGTCTCGGATTGCAAAAGACGGGAGGACGAGTACCGCTGT----- 467
95 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 114

Db 468 -----GTCCGGTGGTGGTTCAGAAATGCCGTGCTCCAGGTGTTC 506
QY 115 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla 134
Db 507 ACAGCTGCT-----TCGTGGAAGACCATGTGCTCCGATGACTGGAAGGGTCACTACGCA 560
QY 135 GluThrAlaCysArgGlnMetGlyTyrSerSer----- 145
Db 561 AATGTTGCTGTGCCCAACTGGGTTTCCCAAGCTATGTAGTTCAGATAACCTCAGAGTG 620
QY 146 -----LysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAspVal 162
Db 621 AGCTCGCTGAGGGGCGAGTTCCGGGAGGAGTTTGTGTCCATCGATCACCTCTTGCCAGAT 680
QY 163 ValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCysLeu 182
Db 681 GACAAGGTGACTGCATTACACCACTCAGTATATGTGAGGGAGGGA-----TGTGCC 731
QY 183 SerGlySerLeuValSerLeuHisCysLeuAlaCysGly---LysSerLeuLysThrPro 201
Db 732 TCTGGCCACGTGGTTACCTTGCAGTGCACAGCCTGTGGTTCATAGAGGGGCTACAGTCA 791
QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
Db 792 CGCATCGTGGGTGGAACATGTCTTGTCTCTCGAGTGGCCCTGGCAGGCCAGCCTTCAG 851
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
Db 852 TTCCAGGGCTACCACTGTGCGGGGCTCTGTTCATCAGCCCTGTGGATCATCACTGCT 911
QY 242 AlaHisCysPheArgLysHisThrAspValPhe-----AsnTyrLysValArgAla 258
Db 912 GCACACTGTGTT-----TATGACTTGTACCTCCCAAGTTCATGGACCATCCAGGTG 962
QY 259 GlySerAspLysLeu-----GlySerPheProSerLeuAlaValAlaLysIleIleIle 276
Db 963 GGTCTAGTTCCCTGTTGGACAATCCAGCCCATCCCACTTGGTGGAGAAGATTGTT--- 1019
QY 277 IleGluPheAsnProMetTyr---ProLys-----AspAsnAspIleAlaLeuMetLys 293
Db 1020 -----TACCACAGCAAGTACAGCAAAAGAGGCTGGGCAATGACATCGCCCTTATGAAG 1073
QY 294 LeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAsp 313
Db 1074 CTGCCCGGGCACTCACGTTCAATGAAATGATCCAGCCTGTGTGCTGCCCAACTCTGAA 1133
QY 314 GluGluLeuThrProAlaThrProLeuTyrIleIleGlyTrpGlyPheThrLysGlnAsn 333
Db 1134 GAGAACTTCCCGGATGAAAAGTGTGCTGGACGTGAGGATGGGGGCCACA---GAGGAT 1190
QY 334 GlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArg 353
Db 1191 GGAGGTGACGCTCCCTGTCTCTGAACCAACGCGGCGCTCCCTTTGATTTCACAAAGAT 1250
QY 354 -CysAsnAlaAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIle 373
Db 1251 CTGCAACCAACAGGACGCTGTACGTTGGCATCATCTCCCTCCATGCTCTGCGCGGCTA 1310
QY 373 eProGluGlyGlyValAspThr--CysGlnGlyAspSerGlyGlyProLeuMetTyrGln 392
Db 1311 CCTGACGGTGGCGTGTGAACAGCTGCCAGGGGACAGCGGGGGCCCCCTGTGTGTCAA 1370
QY 393 SerAspGln---TyrHisValValGlyIleValSerTrpGlyTyrGlyCysGlyPro 411
Db 1371 GAGAGGAGCTGTGGAAGTTAGTGGAGCGACCACTTGGCATCGGCTCGGCACACGTG 1430
QY 412 SerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTyrIleTyrAsnValTrp 431
Db 1431 AACAAAGCCTGGGTGTACACCCGTGTACCTCTCTGACTGGATCCACGAGCAGATG 1490
QY 432 LysAlaGluLeu 435
Db 1491 GAGAGAGACCTA 1502

RESULT 10
 S-09-510-738A-188
 Sequence 188, Application US/09510738A
 Patent No. 6268165
 GENERAL INFORMATION:
 APPLICANT: O'Brien, Timothy J.
 TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
 TITLE OF INVENTION: Ovarian Cancer
 FILE REFERENCE: D6223CIP-A
 CURRENT APPLICATION NUMBER: US/09/510,738A
 CURRENT FILING DATE: 2000-02-22
 PRIOR APPLICATION NUMBER: 09/039,211
 PRIOR FILING DATE: 03-14-1998
 NUMBER OF SEQ ID NOS: 188
 SEQ ID NO 188
 LENGTH: 1783
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: full length cDNA of hepsin
 JS-09-510-738A-188
 Alignment Scores:
 Pred. No.: 1.36e-53 Length: 1783
 Score: 584.50 Matches: 140
 Percent Similarity: 45.05% Conservative: 60
 Best Local Similarity: 31.53% Mismatches: 161
 Query Match: 24.96% Indels: 83
 DB: 3 Gaps: 11
 JS-09-607-745-2 (1-435) x US-09-510-738A-188 (1-1783)
 QY 19 LysProArgIleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeu 38
 Db 288 AGACCAAGTGGCAGCTCTCACT-----GCGGGACCCCTGCTACTTCTGACAGCC 338
 QY 39 LeuSerLeuAlaSerIleIleIleValValIleValIleLysValIleLeuAspLysTyr 58
 Db 339 ATCGGGGGCGGCATCCTGGGCCATTGTGGCTGTTCTCTCAGG----- 380
 QY 59 TyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGlu 78
 Db 380 -----AGTGACCAGGAG----- 380
 QY 79 LeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyPro 98
 Db 381 -----AGTGACCAGGAG-----CCG 395
 QY 99 AlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThr 118
 Db 396 CTGTACCCAGTGCAGGTCTGCTCTCGGACGCTCGGCTCATGCTTTTGACAAAGACGGAA 455
 QY 119 GlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCys 138
 Db 456 GGGACGTGGCGGTGCTGTGCTCTCGCGCTCCACGCCAGGGTAGCCGGACTCAGCTGC 515
 QY 139 ArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGln 158
 Db 516 GAGGAGATGGGCTTC-----CTCAGGGCACTG-----ACCCACTCC 551
 QY 159 AspLeuAspValValGluIleThrGluAsn----- 168
 Db 552 GAGCTGGACGTGCGAACGGCGGGGCCAATGGCACGTCGGGCTTCTTGTGTGGACGAG 611
 QY 169 -----SerGlnGluLeuArgMetArgAsnSerSerGlyProCysLeuSer 183
 Db 612 GGGAGGTGCCCCACCCAGAGGCTGTGGAGGTATCTCGTGTGTGATGCCCCAGA 671
 QY 184 GlySerLeuValSerLeuHisCysLeuAlaCysGly---LysSerLeuLysThrProArg 202
 Db 672 GGCCGTTTCTGGCCGCCATCTGCCAAGACTGTGGCCGCGCAGGAAGCTGCCCGTGGACCGC 731

QY 203 ValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGlnTyr 222
 Db 732 ATCGTGGAGCGCGGACACACAGCTTGGGCCGCTGGCCGTGGAGTGGAGTTCGCTAT 791
 QY 223 AspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAlaAla 242
 Db 792 GATGGAGCACACCTCTGTGGGGATCCCTGCTCTCCGGGACTGGGTGTGACAGCCGCC 851
 QY 243 HisCysPheArgLysHisThrAspValPheAsn---TrpLysValArgAlaGlySerAsp 261
 Db 852 CACTGCTTCCCGAGCGGAACCGGCTCCTGTCCGATGGCGAGTGTTCGGGTGCCGTG 911
 QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
 Db 912 GCCCAGGCTCTCCACCGTCTGCAGCTGGGGTGCAGGCTGTGTCTACCACGGGGC 971
 QY 282 MetTyrPro-----LysAspAsnAspIleAlaLeuMetLys 293
 Db 972 TATCTTCCCTTTCGGGACCCCAACAGCGAGGAGAACAGCAACGATATTGCCCTGGTCCAC 1031
 QY 294 LeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAsp 313
 Db 1032 CTCTCCAGTCCCTGCCCCCTCACAGATACATCCAGCTGTGTGCTCCAGCTGCCGGC 1091
 QY 314 GluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsn 333
 Db 1092 CAGGCCCTGTGGATGGCAAGATCTGTACCGTGACGGCTGGGGCAACACAG---CAGTAC 1148
 QY 334 GlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArg 353
 Db 1149 TATGCCCAACAGCGCGGGTACTCCAGGAGCTCGAGTCCCCATATCAGCAATGATGC 1208
 QY 354 CysAsnAlaAspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIle 373
 Db 1209 TGCAATGGCGCTGACTTCTATGGAACCCAGATCAAGCCCAAGATGTCTGTGTGGCTAC 1268
 QY 374 ProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGln--- 392
 Db 1269 CCCGAGGTGGCATTGATGCTGCCAGGGGACAGCGGTGCTCCCTTGTGTGTGAGGAC 1328
 QY 393 -----SerAspGlnTrpHisValValGlyIleValSerTrpGlyTyrGlyCys 408
 Db 1329 AGCATCTCTCGGACGCCACGTTGGCGGCTGTGTGGCATTGTGAGTTGGGGCACTGGCTGT 1388
 QY 409 GlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyr 428
 Db 1389 GCCTGGGCCCAGAACGAGCGGTCTACACCAAAAGTCAGTGACTTCCGGGAGTGGATCTTC 1448
 QY 429 AsnValTrpLys 432
 Db 1449 CAGGCCATAAG 1460
 RESULT 11
 US-09-861-966-188
 Sequence 188, Application US/09861966
 Patent No. 6518028
 GENERAL INFORMATION:
 APPLICANT: O'Brien, Timothy J.
 TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
 TITLE OF INVENTION: Ovarian Cancer
 FILE REFERENCE: D6223CIP-A/Div
 CURRENT APPLICATION NUMBER: US/09/861,966
 CURRENT FILING DATE: 2001-05-21
 PRIOR APPLICATION NUMBER: 09/510,738
 PRIOR FILING DATE: 2000-02-22
 NUMBER OF SEQ ID NOS: 188
 SEQ ID NO 188
 LENGTH: 1783
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: full length cDNA of hepsin
 US-09-861-966-188

Alignment Scores:		1.36e-53		Length: 1783	
ed. No.:		584.50		Matches: 140	
Percent Similarity:		45.05%		Conservative: 60	
st Local Similarity:		31.53%		Mismatches: 161	
ery Match:		24.96%		Indels: 83	
:		4		Gaps: 11	
-09-607-745-2 (1-435) x US-09-861-966-188 (1-1783)					
Qy	19	LysProArgIleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeu	38		
Db	288	AGACCAAGTGGCAGCTCTCACT-----GCGGGACCCCTGCTACTTCTGACAGCC	338		
Qy	39	LeuSerLeuAlaSerIleIleValValLeuIleLysValIleLeuAspLysTyr	58		
Db	339	ATCGGGCGGCATCCTGGGCCATTGTGGCTGTCTCCTCAGG-----	380		
Qy	59	TyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGlu	78		
Db	380	-----	380		
Qy	79	LeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyPro	98		
Db	381	-----AGTGACCAGGAG-----	395		
Qy	99	AlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThr	118		
Db	396	CTGTACCCAGTGCAGGTGAGCTCTGCGGACGCTCGGCTCATGGTCTTTGACAAGACGGAA	455		
Qy	119	GlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCys	138		
Db	456	GGGACGTGGCGGTGCTGTCTCTCGCGCTCCAACGCCAGGGTAGCCGAGCTCAGTGC	515		
Qy	139	ArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGln	158		
Db	516	GAGGAGATGGGCTTC-----CTCAGGGCACTG-----ACCCACTCC	551		
Qy	159	AspLeuAspValValGluIleThrGluAsn-----	168		
Db	552	GAGCTGGACGTGGCAACGGCGGGCGCCATGGCACGTCGGGCTTCTTCTGTGTGGACGAG	611		
Qy	169	-----SerGlnGluLeuArgMetArgAsnSerSerGlyProCysLeuSer	183		
Db	612	GGGAGGCTGCCCCACACCCAGAGGCTGCTGGAGGTTCATCTCCGTGTGTGATGCCCCAGA	671		
Qy	184	GlySerLeuValSerLeuHisCysLeuAlaCysGly---LysSerLeuLysThrProArg	202		
Db	672	GGCCGTTTCTTGCCGCCATCTGCCAAGACTGTGGCCGCGAGGAAGTGCCCGTGGACCCGC	731		
Qy	203	ValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGlnTyr	222		
Db	732	ATCGTGGGAGCGCGGACACACAGCTTGGCGCGTGGCGGTGGCAAGTTCAGCCTTCGCTAT	791		
Qy	223	AspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAlaAla	242		
Db	792	GATGGAGCACACCTCTGTGGGGGATCCCTGCTCTCCGGGACTGGGTGCTGACAGCCGCC	851		
Qy	243	HisCysPheArgLysHisThrAspValPheAsn---TrpLysValArgAlaGlySerAsp	261		
Db	852	CACGTCTCCCGAGCGGAACCGGGTCTCTCCCGATGGCGAGTGTTCGGGTGGCCGTG	911		
Qy	262	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro	281		
Db	912	GCCAGGCCTCTCCCCACCGTCTGCAGCTGGGGTGCAGGCTGTGTGTCTACACGGGGGC	971		
Qy	282	MetTyrPro-----LysAspAsnAspIleAlaLeuMetLys	293		
Db	972	TATCTTCCCTTTCGGGACCCCAACAGCGAGGAGAACAGCATATATGCTGCTGCCAC	1031		
Qy	294	LeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPheAsp	313		

Qy	314	GluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsn	333		
Db	1092	CAGGCCCTGGTGGATGGCAAGATCTGTACCTGACGGGCTGGGGCAACACG---CAGTAC	1148		
Qy	334	GlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArg	353		
Db	1149	TATGGCCACACAGGCCGGGTACTCCAGGAGGCTCGAGTCCCCATAATCAGCAATGATGTC	1208		
Qy	354	CysAsnAlaAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIle	373		
Db	1209	TGCAATGGCGCTGACTTCTATGGAAACACAGATCAAGCCCAAGATGTTCTGTGTGGCTAC	1268		
Qy	374	ProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGln---	392		
Db	1269	CCGAGGGTGGCATTGATGCCTGCCAGGGCGACAGCGGTGTCCTTTGTGTGTGAGGAC	1328		
Qy	393	-----SerAspGlnTrpHisValValGlyIleValSerTrpGlyTyrGlyCys	408		
Db	1329	AGCATCTCTCGACGCCACGTTGGCGGCTGTGTGGCATTTGAGTTGGGCACCTGGCTGT	1388		
Qy	409	GlyGlyProSerThrProGlyValThrLysValSerAlaTyrLeuAsnTrpIleTyr	428		
Db	1389	GCCCTGGCCCAAGCCAGGCGTCTACACCAAGTCAGTGACTCCGGGAGTGGATCTTC	1448		
Qy	429	AsnValTrpLys	432		
Db	1449	CAGGCCATAAAG	1460		
RESULT 12					
US-09-742-703-3					
; Sequence 3, Application US/09742703					
; Patent No. 6423543					
; GENERAL INFORMATION:					
; APPLICANT: Patrick Allen Marcotte					
; APPLICANT: Lex M. Cowser					
; TITLE OF INVENTION: ANTISENSE MODULATION OF HEP SIN EXPRESSION					
; FILE REFERENCE: RTS-0090					
; CURRENT APPLICATION NUMBER: US/09/742,703					
; CURRENT FILING DATE: 2000-12-20					
; NUMBER OF SEQ ID NOS: 49					
; SEQ ID NO 3					
; LENGTH: 2363					
; TYPE: DNA					
; ORGANISM: Homo sapiens					
; FEATURE:					
; NAME/KEY: CDS					
; LOCATION: (826)..(2079)					
US-09-742-703-3					
Alignment Scores:					
Pred. No.:		2.11e-53		Length: 2363	
Score:		584.50		Matches: 140	
Percent Similarity:		45.05%		Conservative: 60	
Best Local Similarity:		31.53%		Mismatches: 161	
Query Match:		24.96%		Indels: 83	
DB:		4		Gaps: 11	
US-09-607-745-2 (1-435) x US-09-742-703-3 (1-2363)					
Qy	19	LysProArgIleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeu	38		
Db	868	AGACCAAGTGGCAGCTCTCACT-----GCGGGACCCCTGCTACTTCTGACAGCC	918		
Qy	39	LeuSerLeuAlaSerIleIleValValLeuIleLysValIleLeuAspLysTyr	58		
Db	919	ATCGGGCGGCATCTCTGGCCATTGTGGCTGTCTCCTCAGG-----	960		
Qy	59	TyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGlu	78		
Db	960	-----	960		

Db	1032	CTCTCCAGTCCCTGCCCTCACAGATAATCCAGCCTGTGTGCCTCCAGCTGCCGCGC	1091
Qy	314	GluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsn	333
Db	1092	CAGGCCCTGGTGGATGGCAAGATCTGTACCGTGACGGGCTGGGGCAACACG--CAGTAC	1148
Qy	334	GlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArg	353
Db	1149	TATGGCCAAACAGGCCGGGTACTCCAGGAGGCTCGAGTCCCCATAATCAGCAATGATGTC	1208
Qy	354	CysAsnAlaAspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIle	373
Db	1209	TGCAATGGCGCTGACTTCTATGGAACACAGATCAAGCCCAAGATGTTCTGTGTGGCTAC	1268
Qy	374	ProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGln---	392
Db	1269	CCCGAGGTGGCATTGATGCTCTGCCAGGGCGACAGCGTGGTCCCTTTGTGTGTGAGGAC	1328
Qy	393	-----SerAspGlnTrpHisValValGlyIleValSerTrpGlyTyrGlyCys	408
Db	1329	AGCATCTCTCGGACGCCACGTTGGCGGCTGTGTGGCATTTGTAGTTGGGGCACTGGCTGT	1388
Qy	409	GlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyr	428
Db	1389	GCCCTGGCCAGAACGCCAGCGCTCTACACCAAGTCAGTGACTTCGGGGAGTGGATCTTC	1448
Qy	429	AsnValTrpLys	432
Db	1449	CAGGCCATAAAG	1460
RESULT 12			
US-09-742-703-3			
; Sequence 3, Application US/09742703			
; Patent No. 6423543			
; GENERAL INFORMATION:			
; APPLICANT: Patrick Allen Marcotte			
; APPLICANT: Lex M. Cowser			
; TITLE OF INVENTION: ANTISENSE MODULATION OF HEP SIN EXPRESSION			
; FILE REFERENCE: RTS-0090			
; CURRENT APPLICATION NUMBER: US/09/742,703			
; CURRENT FILING DATE: 2000-12-20			
; NUMBER OF SEQ ID NOS: 49			
; SEQ ID NO 3			
; LENGTH: 2363			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (826)..(2079)			
US-09-742-703-3			
Alignment Scores:			
Pred. No.:	2.11e-53	Length:	2363
Score:	584.50	Matches:	140
Percent Similarity:	45.05%	Conservative:	60
Best Local Similarity:	31.53%	Mismatches:	161
Query Match:	24.96%	Indels:	83
DB:	4	Gaps:	11
US-09-607-745-2 (1-435) x US-09-742-703-3 (1-2363)			
Qy	19	LysProArgIleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeu	38
Db	868	AGACCAAGGTGGCAGCTCTCACT-----GCGGGACCCCTGCTACTTCTGACAGCC	918
Qy	39	LeuSerLeuAlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyr	58
Db	919	ATCGGGCGGCATCCTGGGCCATTGTGGCTGTCTCCTCAGG-----	960
Qy	59	TyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGlu	78
Db	960	-----	960

79 LeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyPro 98
961 -----AGTGACCAGGAG-----CCG 975
99 AlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThr 118
976 CTGTACCCAGTGAGGTCTGCGGACGCTCGGCTCATGGTCTTTGACAAGACGGAA 1035
119 GlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCys 138
1036 GGGACGTGGCGGCTGCTGCTCCTCGCGCTCCACGCCAGGGTAGCCGACTCAGCTGC 1095
139 ArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGln 158
1096 GAGGAGATGGGCTTC-----CTCAGGGCACTG-----ACCCACTCC 1131
159 AspLeuAspValValGluIleThrGluAsn-----168
1132 GAGCTGGACGTGGGAACGGCGGGCGCCAATGGCAGCTCGGGCTTCTTCTGTGTGGACGAG 1191
169 -----SerGlnGluLeuArgMetArgAsnSerSerGlyProCysLeuSer 183
1192 GGGAGGCTGCCACACCCAGAGGCTGCTGGAGTCACTCCGTGTGTAATGCCCCAGA 1251
184 GlySerLeuValSerLeuHisCysLeuAlaCysGly---LysSerLeuLysThrProArg 202
1252 GCGCGTTCTTGGCGCGCATCTGCCAAGACTGTGGCGCGCAGGAAGCTGCCGTGGACCGC 1311
203 ValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIleGlnTyr 222
1312 ATCGTGGAGGCGGGACACAGCTTGGCGCGGTGGCGGTGGCAAGTCAAGCTTCGCTAT 1371
223 AspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAlaAla 242
1372 GATGGAGCACACCTCTGTGGGGGATCCCTGCTCTCGGGGACTGGGTGCTGACAGCCGCC 1431
243 HisCysPheArgLysHisThrAspValPheAsn---TrpLysValArgAlaGlySerAsp 261
1432 CACTGCTTCCCGAGCGGGAACCGGGTCTCTGCCGATGGCGAGTGTTCGCGGTGCCGTG 1491
262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
1492 GCCAGGCTCTCCCCACGGTCTGACGCTGGGGGTGAGGCTGTGGTCTACACGGGGGC 1551
282 MetTyrPro-----LysAspAsnAspIleAlaLeuMetLys 293
1552 TATCTTCCCTTTCGGGACCCCAACAGCAGGAGGAGAACAGATATTCGCTGCTCCAC 1611
294 LeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAsp 313
1612 CTCTCCAGTCCCTGCCCCCTCACAGATACATCCAGCTGTGTGCTCCAGTCCCGGC 1671
314 GluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsn 333
1672 CAGGCCCTGGTGGATGGCAAGATCTGTACCGTGACGGGTGGGGCAACACG---CAGTAC 1728
334 GlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArg 353
1729 TATGGCCAAACAGCGCGGGTACTCCAGGAGGCTCGAGTCCCCATAATCAACAATGATGC 1788
354 CysAsnAlaAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIle 373
1789 TGCAATGGCGCTGACTTCTATGGAACACAGATCAAGCCCAAGATGTTCTGTGTGGCTAC 1848
374 ProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGln--- 392
1849 CCCGAGGTGGCATTGATGCTTGCAGGGCGACAGCGGTGGTCCCTTTGTGTGTGAGGAC 1908
393 -----SerAspGlnTrpHisValValGlyIleValSerTrpTrpGlyCys 408
1909 AGCATCTCTCGGACGCCACGTGGCGGCTGTGTGGCATGTGAGTTGGGCACTGGCTGT 1968
409 GlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyr 428

Db 1969 GCCCTGGCCCAAGCCAGGCGTCTACACCAAGTCACTGACTTCCGGGAGTGGATCTTC 2028
Qy 429 AsnValTrpLys 432
Db 2029 CAGGCCATAAAG 2040
RESULT 13
US-09-820-002-1
; Sequence 1, Application US/09820002
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weiniu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001194
; CURRENT APPLICATION NUMBER: US/09/820,002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1615
; TYPE: DNA
; ORGANISM: HUMAN
US-09-820-002-1
Alignment Scores:
Pred. No.: 3.6e-53 Length: 1615
Score: 580.00 Matches: 132
Percent Similarity: 44.06% Conservative: 57
Best Local Similarity: 30.77% Mismatches: 146
Query Match: 24.77% Indels: 94
DB: 4 Gaps: 9
US-09-607-745-2 (1-435) x US-09-820-002-1 (1-1615)
Qy 19 LysProArgIleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeu 38
Db 220 AGACCAAGGTGGCAGCTCTCACT-----GCGGGACCTGTACTTCTGACAGCC 270
Qy 39 LeuSerLeuAlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyr 58
Db 271 ATCGGGGCGGCATCCTGGGCCATTGTGGCTGTCTCCTCAGG-----312
Qy 59 TyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGlu 78
Db 312 -----312
Qy 79 LeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyPro 98
Db 313 -----AGTGACCAGGAG-----CCG 327
Qy 99 AlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThr 118
Db 328 CTGTACCCAGTGAGGTCTGCTTCTGCGGACGCTCGGCTCATGGTCTTTGACAAGACGGAA 387
Qy 119 GlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCys 138
Db 388 GGGACGTGGCGGCTGTGTCTCTCGCGCTCCAACGCCAGGCTAGCCGACTCAGCTGC 447
Qy 139 ArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGln 158
Db 448 GAGGAGATGGGCTTCCTCAGT-----468
Qy 159 AspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSer 178
Db 468 -----468
Qy 179 GlyProCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGly---LysSer 197

db 469 ---GATGCCCCAGAGCGCGTTCTTGGCCGCCCATCTGCCAAGACTGTGGCCGCAGGAAG 525
198 LeuLysThrProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGln 217
526 CTGCCCGTGGACCGCATCTGTGGGAGGCGCGGACACACAGCTTGGCGCGGTGGCGTGGCAA 585
218 ValSerIleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrp 237
586 GTGAGCCTTCGTATGATGAGACACACCTCTGTGGGGATCCCTGCTCTCCGGGACTGG 645
238 ValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsn---TrpLysVal 256
646 GTGCTGACAGCGCGCCCTGCTTCCCGAGCGGAACCGGTCTCTGCCGATGGCGAGTG 705
257 ArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIle 276
706 TTTGCCGGTGGCGTGGCCAGGCCTCTCCCAACGGTCTGCAGTGGGGTGCAGGCTGTG 765
277 IleGluPheAsnProMetTyrPro-----LysAspAsnAsp 288
766 GTCTACACGGGGGCTATCTTCCCTTTTCGGGACCCCAACAGCGAGGAGAGAACAGCAT 825
289 IleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCys 308
826 ATTGCCCTGGTCCACCTCTCCAGTCCCTGCCCTCCACAGATATACATCCAGCCTGTGTGC 885
309 LeuProPhePheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGly 328
886 CTCCAGCTGCCGGCCAGCCCTGGTGGATGGCAAGATCTGTACCGTGACGGGCTGGGC 945
329 PheThrLysGlnAsnGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnVal 348
946 AACACG---CAGTACTATGSCCAACAGCGCGGGTACTCCAGGAGGCTCGAGTCCCCATA 1002
349 IleAspSerThrArgCysAsnAlaAspAlaTyrGlnGlyGluValThrGluLysMet 368
1003 ATCAGCAATGATGTCTGCAATGGCGTGACTTCTATGGAACACAGATCAAGCCCCAAGATG 1062
369 MetCysAlaGlyIleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyPro 388
1063 TTCTGTGTGGCTACCCGAGGGTGGCATTTGATGCCTGCCAGGCGACAGCGGTGGTCCC 1122
389 LeuMetTyrGln-----SerAspGlnTrpHisValValGlyIleValSer 403
1123 TTTGTGTGTGAGGACAGCATCTCTCGGACGCCACGTTGGCGGTGTGTGGCATTGTGAGT 1182
404 TrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyr 423
1183 TGGGCACTGGCTGTGCCCTGGCCCAAGAGCGGCGCTACACCAAGTCAGTGACTTC 1242
424 LeuAsnTrpIleTyrAsnValTrpLys 432
1243 CGGAGTGGATCTTCCAGGCCATAAG 1269

RESULT 14

US-09-000-846-1
Sequence 1, Application US/09000846
Patent No. 5981830
GENERAL INFORMATION:
APPLICANT: WU, QINGYU
APPLICANT: SADLER, JASPER
TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH
DISRUPTED SERINE PROTEASE GENES
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN, P.C.
STREET: 2200 CLARENDON BLVD. SUITE 1400
CITY: ARLINGTON
STATE: VA
COUNTRY: US
ZIP: 22201
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,846
FILING DATE: 30-DEC-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/866,058
FILING DATE: 30-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: LEBOVITZ, RICHARD M.
REGISTRATION NUMBER: 37,067
REFERENCE/DOCKET NUMBER: BERLX 65P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-243-6333
TELEFAX: 703-243-6410
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1605 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: 35..1282
US-09-000-846-1
Alignment Scores:
Pred. No.: 5.88e-53 Length: 1605
Score: 578.00 Matches: 136
Percent Similarity: 47.43% Conservative: 58
Best Local Similarity: 33.25% Mismatches: 152
Query Match: 24.68% Indels: 63
DB: 2 Gaps: 11
US-09-607-745-2 (1-435) x US-09-000-846-1 (1-1605)
Qy 62 CysGlyGlnProLeuHisPheIleProArgLysGln-----LeuCysAspGlyGluLeu 79
Db 97 TGTGGGTACCTGCT-----GTTCTGACAGGCATTGGGGCCGCTCTGGGCCATTGT 150
Qy 80 AspCysProLeuGlyGlu-AspGluGluHisCysValLysSerPheProGluGlyProAl 99
Db 151 GACCATCTACTGACAGAGTGACCAGGAG-----CCACT 183
Qy 99 aValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGl 119
Db 184 GTACCAAGTGCAGCTCAGTCCAGGGGACTCAGCGCTTGGCGTGTGTTGACAAAGACGGAGG 243
Qy 119 yAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysAr 139
Db 244 AACGTGGAGGCTACTGTCTCTCCTCAGCTCCAATGCCAGGGTGGCAGGGCTCGGCTGTGA 303
Qy 139 gGlnMetGlyTyr-----143
Db 304 GGAGATGGGCTTTCTCAGGGCTCTGGCGCACTCGGAGCTGGATGTCCGCACTCGGGCGC 363
Qy 144 -SerSerLysProThrPheArgAlaValGluIleGly-----ProAspGlnAs 159
Db 364 CAACGGCACATCGGGCTTCTTTTGGTGGACGAGGCGGAGCTCCGCTGGCTCAGAGTT 423
Qy 159 pLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGl 179
Db 424 GCTGGATGTCTCTGTATGTGAC-----448
Qy 179 yProCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGly---LysSerLe 198

```
b      449  ---TGTCCTAGAGCGCGATTCCTGACTGCCACCTGCCAAGACTGTGGCCGACGAAGCT 504
      y      198  uLysThrProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnVa 218
      b      505  GCCGGTGGACCGCATTTGGGGGGCCAGGACAGCAGTCTGGGAAGGTGGCCGTGGCAGGT 564
      y      218  lSerIleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpVa 238
      b      565  CAGCCTGCGTTATGATGGACCCACCTCTGTGGGGGTCCCTGTGTGGGACTGGGT 624
      y      238  lLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsn---TrpLysValAr 257
      b      625  GCTGACTGCTGCACATTGCTTCCAGAGCGGAACCGGTCCTGTCTCGGTGGCGAGTATT 684
      y      257  gAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleI 277
      b      685  TGCTGGTGTGTAGCCCGGACCTCACCCCATGCTGTGCAACTGGGGGTTCAGGCTGTGAT 744
      y      277  eGluPheAsnProMetTyrPro-----LysAspAsnAspI 289
      b      745  CTATCATGGGGGCTACCTTCCTTCGAGACCCCTACTATCGACGAAACACCAATGACAT 804
      y      289  eAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLe 309
      b      805  TGCCTTGGTCCACCTCTCTAGTCCCTGCCTCTCACAGAATACATCCAGCCAGTGTGTCT 864
      y      309  uProPhePheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPh 329
      b      865  CCCTGTGCGGGACAGCGCCCTGTGTGGATGGCAAGTCTGTACTGTGACCGGTGGGTAA 924
      y      329  eThrLysGlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValI 349
      b      925  CACA---CAGTTCTATGGCCAAACAGGCTATGTTGCTCCAAAGAGGCCCGGTGCCATCAT 981
      y      349  eAspSerThrArgCysAsnAlaAspAspAlaTyrGlnGlyGluValThrGluLysMetMe 369
      b      982  AAGCAACGAAGTTTGCAACAGCCCGGACTTCTACGGGAATCAGATCAAGCCCAAGATGTT 1041
      y      369  tCysAlaGlyIleProGluGlyValAspThrCysGlnGlyAspSerGlyGlyProLe 389
      b      1042  CTGTGTGCTATCCTGAGGGTGGCATTTGATGCGTGCCAGGGCGACAGTGGAGGCCCTTT 1101
      y      389  uMetTyrGln-----SerAspGlnTrpHisValValGlyIleValSerTr 404
      b      1102  TGTGTGTGAAGACAGCATCTCTGGGACATCAAGTGGCGGTATGTGGCATTTGAAGCTG 1161
      y      404  pGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyrLe 424
      b      1162  GGGTACGGGCTGTGCTTTGGCCCGGAAGCCAGGAGTGTACACCAAAAGTCACTGACTCCG 1221
      y      424  uAsnTrpIleTyrAsnValTrpLys 432
      b      1222  GGAGTGGATCTTCAAGGCCATAAAG 1246
```

```
RESULT 15
US-09-370-838-80
Sequence 80, Application US/09370838
Patent No. 644425
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Mohamath, Roadoh
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.475C1
CURRENT APPLICATION NUMBER: US/09/370,838
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/285,323
EARLIER FILING DATE: 1999-04-02
NUMBER OF SEQ ID NOS: 289
SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 80
; LENGTH: 1460
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-370-838-80

Alignment Scores:
Pred. No.: 8.37e-53 Length: 1460
Score: 576.00 Matches: 145
Percent Similarity: 50.88% Conservative: 86
Best Local Similarity: 31.94% Mismatches: 161
Query Match: 24.59% Indels: 62
DB: 4 Gaps: 16

US-09-607-745-2 (1-435) x US-09-370-838-80 (1-1460)

Qy      5 SerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArgIleProMet 24
Db      19 GCAGAAAAAAGAACCTCTTCAATTAAGGATTAAATGTATAGGCCAGCACGTTGA---ACT 75
Qy      25 GluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeuAlaSerIle 44
Db      76 TCGACTTCAAGATTCTGAAT---CCATATGTAGTATGTTTTCATTTCTGTCGCGAGGGTA 132
Qy      45 IleIleValValLeuIleLysValIleLysValIleLeuAspLysTyrTyrPheLeuCysGlyGln 64
Db      133 GTGATCCTGGCAGTCACCATAGCTCTACTTGT-----TACTTTTGTAGCTTTTGTAT 183
Qy      65 ProLeuHisPheIleProArgLys-----GlnLeuCysAspGlyGluLeuAspCysPro 82
Db      184 CAAAAATCTTACTTTTATAGGAGCAGTTTTCAACTCCTAAATGTTGAATATAATAGTCAG 243
Qy      83 Leu-----GlyGluAspGluGluHisCys 90
Db      244 TTAATTCCACGCTACACAGGAATACAGGACTTTGAGTGAAGAAATTGAATCTCTGATT 303
Qy      91 ValLysSerPheProGluGly-----ProAlaValAlaValArg 103
Db      304 ACTAAACATTTCAAGAAATCAAAATTTAAGAAATCAGTTTCATCAGAGCTCATGTTGCCAAA 363
Qy      104 LeuSerLysAsp-----ArgSerThrLeuGlnValLeuAspSerAlaThrGly 119
Db      364 CTGAGGCAAGATGGTAGTGTGTGAGCGGATGTTGTTCATGAAATTTCAATTCACCTAGA 423
Qy      120 AsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArg 139
Db      424 AATAACAATGGAGCATCAATG-----AAAAGCAGAAATTGAGTCTGTTTACGA 471
Qy      140 GlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAsp 159
Db      472 CAAATGCTGAATAACTCT-----GGAAACCTGGAATAAAACCT-----TCA 513
Qy      160 LeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsn-----Ser 177
Db      514 ACTGAGATAACATCACTTACTGACCGCTGCAGCAAAATGGCTTATTATGAATGTGGG 573
Qy      178 SerGlyProCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSer 197
Db      574 GCCGTCCAGACCTAATA-----GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 594
Qy      198 LeuLysThrProArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGln 217
Db      595 TTGTCTGAGCAGAGAATCCTTGGAGGCACTGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 654
Qy      218 ValSerIleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrp 237
Db      655 GTCAGTCTGGGCTCAATAATATGCCACCTGTCAGGAGCAGCCTGATCAATAACATGTGG 714
Qy      238 ValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArg 257
Db      715 ATCCTGACAGCAGCTCACTGTCTCAGAGCAACTCTAATCCTCGTACTGGATGCCACG 774
Qy      258 AlaGlySerAspLysLeuGlySerPheProSerLeu-----AlaValAlaLysIleIle 275
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model
run on: November 23, 2003, 07:54:55 ; Search time 467.909 Seconds
(without alignments)
3038.343 Million cell updates/sec

itle: US-09-607-745-2
erfect score: 2342
equense: 1 MDPDSQPLNSLDVKPLKRP.....VYTKVSAYLNWIYNVWKAEI 435

coring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

sarched: 2169961 seqs, 1634102185 residues

otal number of hits satisfying chosen parameters: 4339922

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

ommand line parameters:
MODEL=frame+p2n.model -DEV=xlh
2=/cgn2_1/USPTO_spool/US09607745/runat 21112003 144347 22270/app_query.fasta_1.1038
DB=Published_Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
FRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0
MAXLEN=2000000000 -USER=US09607745 @CGN 1 1 424 @runat 21112003 144347 22270
NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match Length	ID	Description
1	2338	99.8	2038 12	US-10-180-719-18 Sequence 18, Appl

2	2337	99.8	2081	9	US-09-851-588-7	Sequence 7, Appli
3	2337	99.8	2307	14	US-10-097-340-317	Sequence 317, App
4	2337	99.8	2307	14	US-10-171-311-217	Sequence 217, App
5	2329	99.4	2165	12	US-10-101-510-634	Sequence 634, App
6	2324	99.2	2079	9	US-09-851-588-5	Sequence 5, Appli
7	2324	99.2	2079	11	US-09-776-191-71	Sequence 71, Appl
8	2324	99.2	2079	14	US-10-264-820-22	Sequence 22, Appl
9	2324	99.2	2079	14	US-10-254-289-1	Sequence 1, Appli
10	2319	99.0	2137	11	US-09-776-191-3	Sequence 3, Appli
11	2297.5	98.1	2063	11	US-09-888-257A-2	Sequence 2, Appli
12	2297.5	98.1	2063	11	US-09-946-374-274	Sequence 274, App
13	2297.5	98.1	2063	12	US-10-015-387A-274	Sequence 274, App
14	2297.5	98.1	2063	12	US-10-063-735-111	Sequence 111, App
15	2297.5	98.1	2063	12	US-10-006-130A-274	Sequence 274, App
16	2297.5	98.1	2063	12	US-10-199-672-329	Sequence 329, App
17	2297.5	98.1	2063	12	US-10-006-172A-274	Sequence 274, App
18	2297.5	98.1	2063	12	US-10-187-749-329	Sequence 329, App
19	2297.5	98.1	2063	12	US-10-194-457-329	Sequence 329, App
20	2297.5	98.1	2063	12	US-10-184-642-329	Sequence 329, App
21	2297.5	98.1	2063	12	US-10-196-747-329	Sequence 329, App
22	2297.5	98.1	2063	12	US-10-015-392A-274	Sequence 274, App
23	2297.5	98.1	2063	12	US-10-017-253A-274	Sequence 274, App
24	2297.5	98.1	2063	12	US-10-173-689-329	Sequence 329, App
25	2297.5	98.1	2063	12	US-10-173-690-329	Sequence 329, App
26	2297.5	98.1	2063	12	US-10-173-691-329	Sequence 329, App
27	2297.5	98.1	2063	12	US-10-173-692-329	Sequence 329, App
28	2297.5	98.1	2063	12	US-10-173-694-329	Sequence 329, App
29	2297.5	98.1	2063	12	US-10-173-698-329	Sequence 329, App
30	2297.5	98.1	2063	12	US-10-173-699-329	Sequence 329, App
31	2297.5	98.1	2063	12	US-10-173-707-329	Sequence 329, App
32	2297.5	98.1	2063	12	US-10-174-569-329	Sequence 329, App
33	2297.5	98.1	2063	12	US-10-174-583-329	Sequence 329, App
34	2297.5	98.1	2063	12	US-10-174-587-329	Sequence 329, App
35	2297.5	98.1	2063	12	US-10-174-589-329	Sequence 329, App
36	2297.5	98.1	2063	12	US-10-174-591-329	Sequence 329, App
37	2297.5	98.1	2063	12	US-10-175-736-329	Sequence 329, App
38	2297.5	98.1	2063	12	US-10-175-742-329	Sequence 329, App
39	2297.5	98.1	2063	12	US-10-175-744-329	Sequence 329, App
40	2297.5	98.1	2063	12	US-10-175-745-329	Sequence 329, App
41	2297.5	98.1	2063	12	US-10-175-748-329	Sequence 329, App
42	2297.5	98.1	2063	12	US-10-175-751-329	Sequence 329, App
43	2297.5	98.1	2063	12	US-10-175-754-329	Sequence 329, App
44	2297.5	98.1	2063	12	US-10-176-480-329	Sequence 329, App
45	2297.5	98.1	2063	12	US-10-176-489-329	Sequence 329, App

ALIGNMENTS

RESULT 1
US-10-180-719-18
; Sequence 18, Application US/10180719
; Publication No. US20030166246A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/180,719
FILING DATE: 25-Jun-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271
FILING DATE: 16-Jan-1998
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNNOT13
CLONE: 1337018

SEQUENCE DESCRIPTION: SEQ ID NO: 18:

S-10-180-719-18

Alignment Scores:

red. No.:	6.19e-283	Length:	2038
core:	2338.00	Matches:	434
percent Similarity:	100.00%	Conservative:	1
est Local Similarity:	99.77%	Mismatches:	0
very Match:	99.83%	Indels:	0
B:	12	Gaps:	0

S-09-607-745-2 (1-435) x US-10-180-719-18 (1-2038)

Y	1	MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro	20
b	200	ATGGATCCTGACAGTGCATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAACCC	259
Y	21	ArgIleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSer	40
b	260	CGTATCCCCATGGAGACCTTCAGAAAGGTGGGATCCCCATCATCATAGCACTACTGAGC	319
Y	41	LeuAlaSerIleIleIleValValValIleLysValIleLeuAspLysTyrTyrPhe	60
b	320	CTGGCGAGTATCATATTGTTGTTCTCTCATCAAGGTGATTCTGGATAAATACTACTTC	379
Y	61	LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp	80
b	380	CTCTGGGGGAGCCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGAC	439
Y	81	CysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaVal	100
b	440	TGTCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTGCAGTG	499
Y	101	AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn	120
b	500	GCAGTCCGCCTCTCCAAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGGAAC	559
Y	121	TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln	140
b	560	TGGTCTCTGCCTGTTTCGACAACCTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAG	619
Y	141	MetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu	160
b	620	ATGGGCTACAGCAGCAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCATCTG	679
Y	161	AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro	180
b	680	GATGTTGTTGAAATCACAGAAAACAGCCAGGAGCTTCGCATGCGGAATCAAGTGGGCC	739
Y	181	CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr	200

Db	740	TGTCCTCAGGCTCCCTGCTCCCTGCACCTGCTTGCCTGTGGGAGAGCCTGAAGACC	799
QY	201	ProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIle	220
Db	800	CCCCGTGTGGTGGTGGGAGGAGCCCTCTGTGGATTCTTGGCCTTGGCAGGTCAGCATC	859
QY	221	GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr	240
Db	860	CAGTACGACAAACAGCACCTCTGTGGAGGAGCATCTGGACCCCCACTGGTCTCTCACG	919
QY	241	AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer	260
Db	920	GCAGCCCACTGCTTCAGGAAACATACCCGATGTGTCAACTGGAAGGTGCGGCAGGCTCA	979
QY	261	AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleLeuGluPheAsn	280
Db	980	GACAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTAATGAATCAAC	1039
QY	281	ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe	300
Db	1040	CCCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTCCCACTCACTTTC	1099
QY	301	SerGlyThrValArgProLysCysLeuProPhePheAspGluGluLeuThrProAlaThr	320
Db	1100	TCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGATGAGGAGCTCACTCCAGCCACC	1159
QY	321	ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle	340
Db	1160	CCACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATA	1219
QY	341	LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyr	360
Db	1220	CTGCTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGTCAATGCAGACGATCGGTAC	1279
QY	361	GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr	380
Db	1280	CAGGGGGAAGTCACCGAGAAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACC	1339
QY	381	CysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGly	400
Db	1340	TGCCAGGTGACAGTGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGC	1399
QY	401	IleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVal	420
Db	1400	ATCGTTAGCTGGGCTATGGCTGCGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTC	1459
QY	421	SerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	435
Db	1460	TCAGCCATATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG	1504

RESULT 2

US-09-851-588-7
; Sequence 7, Application US/09851588
; Patent No. US20020042067A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 2081
; TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (215)..(1528)
OTHER INFORMATION:
S-09-851-588-7

lignment Scores:
red. No.: 8.54e-283 Length: 2081
core: 2337.00 Matches: 434
Percent Similarity: 100.00% Conservative: 0
est Local Similarity: 100.00% Mismatches: 0
very Match: 99.79% Indels: 0
B: 9 Gaps: 0

S-09-607-745-2 (1-435) x US-09-851-588-7 (1-2081)

Y 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
o 224 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAACCCCGT 283
Y 22 IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeu 41
b 284 ATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTAGCCTG 343
Y 42 AlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
o 344 GCGAGTATCATATTGTGGTTGTCTCATCAAGGTGATTCTGGATAAATACTACTTCCTC 403
Y 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
o 404 TGCGGGCAGCCTCTCCACTTTCATCCGAGGAAGCAGCTGTGTGACGGAGAGTGGACTGT 463
Y 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
o 464 CCCTTGGGGAGGAGCAGGAGCAGCTGTCTCAAGAGCTTCCCGAAGGGCCTGCAGTGGCA 523
Y 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
b 524 GTCCGCCCTCTCAAGGACCGATCCCACTGCAGGTGTGGACTCGGCCACAGGGAACTGG 583
Y 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
b 584 TTCTCTGCCTGTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAGATG 643
Y 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
b 644 GGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCAGACCAGGATCTGGAT 703
Y 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
b 704 GTTGTGTAATACAGAAACACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCCCTGT 763
Y 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
b 764 CTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTGCTGTGGGAAGAGCCTGAAGACCCCC 823
Y 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
b 824 CGTGTGGTGGGTGGGAGGAGGCCTCTGTGGATTCTTGGCCCTTGGCAGGTCAGCATCCAG 883
Y 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
b 884 TACGACAAACAGCACCTCTGTGGAGGAGGATCTCTGGACCCCACTGGGTCTCTACGGCA 943
Y 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
b 944 GCCCACTGCTTCAGGAAACATACCAGATGTGTCAACTGGAAGGTGGGGCAGGCTCAGAC 1003
Y 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
b 1004 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAGATCATCATATGATTCAACCCC 1063

QY 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
Db 1064 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCCTCCACTCTTCTCA 1123
QY 302 GlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThrPro 321
Db 1124 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1183
QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
Db 1184 CTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATACTG 1243
QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 361
Db 1244 CTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCAATGCAGACGATGCGTACCAG 1303
QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
Db 1304 GGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGGAAGGGGTGTGGACACCTGC 1363
QY 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
Db 1364 CAGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGGTGGGCATC 1423
QY 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
Db 1424 GTTAGCTGGGGCTATGGCTGCGGGGGCCCGAGCACCCAGGAGTATACCAAGGTCTCA 1483
QY 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
Db 1484 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1525

RESULT 3

US-10-097-340-317
; Sequence 317, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 317

LENGTH: 2307

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)...(2307)

OTHER INFORMATION: n = A,T,C or G

S-10-097-340-317

Lignment Scores:

red. No.: 1e-282 Length: 2307
core: 2337.00 Matches: 434
percent Similarity: 100.00% Conservative: 0
est Local Similarity: 100.00% Mismatches: 0
very Match: 99.79% Indels: 0
B: 14 Gaps: 0

S-09-607-745-2 (1-435) x US-10-097-340-317 (1-2307)

Y 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
b 284 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAACCCCGT 343
Y 22 IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeu 41
b 344 ATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGCCTG 403
Y 42 AlaSerIleIleValValValIleLysValIleLeuAspLysTyrTyrPheLeu 61
b 404 GCGAGTATCATATTGTGTTGTCTCATCAAGGTGATTCTGGATAAATACTACTTCCTC 463
Y 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
b 464 TCGGGGAGCCTCTCCACTTCATCCGAGGAAGCAGCTGTGTGACGGAGAGTGGACTGT 523
Y 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
b 524 CCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTGCAAGTGGCA 583
Y 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
b 584 GTCCGCTCTCTCAAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGGAACCTGG 643
Y 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
b 644 TTCTCTGCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAGATG 703
Y 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
b 704 GGCTACAGCAGCAACCCACTTTCAGAGCTGTGGAGATTGGSCCAGACCAGGATCTGGAT 763
Y 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
b 764 GTTGTGAAATCACAGAAACACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCCCTGT 823
Y 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
b 824 CTCTCAGGCTCCCTGGTCTCCCTGCCTGTCTTGCCTGTGGGAAGAGCCTGAAGACCCCC 883
Y 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
b 884 CGTGTGTTGGTGGGAGGAGGCGCTCTGTGGATTCTTGGCCTTGGCAGGTTCAGCATCCAG 943
Y 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
b 944 TACGACAAACAGCAGCTGTGTGGAGGAGAGATCTCTGGACCCCTCTGGTCTCTCACGGCA 1003
Y 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
b 1004 GCCCACTGCTTCAGGAAACATACCGATGTGTCACTGGAAAGTTCGGGGCAGGCTCAGAC 1063

QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
Db 1064 AAACCTGGGAGCTTCCCATCCCTGGCTGGCCCAAGATCATCATATGAATCAACCCC 1123
QY 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
Db 1124 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTCCCACTCACTTTCTCA 1183
QY 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
Db 1184 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCCCA 1243
QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
Db 1244 CTCTGGATCATTTGGATGGGGCTTTACGAAACAGAAATGGAGGGAAGATGTCTGACATACTG 1303
QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 361
Db 1304 CTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGCAATGCAGACGATGCGTACCAG 1363
QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
Db 1364 GGGGAAGTCAACCGAGAAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGACACCTGC 1423
QY 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
Db 1424 CAGGTGACAGTGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGCATC 1483
QY 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
Db 1484 GTTAGCTGGGGCTATGGCTGGGGGGCCCGACACCCAGGAGTATACACCAAGGTCTCA 1543
QY 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
Db 1544 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1585
RESULT 4
US-10-171-311-217
; Sequence 217, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 217
; LENGTH: 2307
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1691, 1698, 1705, 1708, 1709, 1713, 1717, 1720, 1724, 1728,
; LOCATION: 1733, 1741, 1746, 1748, 1755, 1770, 1774, 1791, 1802, 1821,
; LOCATION: 1838, 1856, 1859, 1864, 1908, 1959, 1997, 2012, 2038, 2143
; OTHER INFORMATION: n = A,T,C or G


```
Y 81 CysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaVal 100
|
b 550 TGTCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGCCCTGCAGTG 609
|
Y 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
|
b 610 GCAGTCCGCCTCTCCAAAGACCGATCCACACTGCAGGTGCTGGACTCGGCCACAGGGAAC 669
|
Y 121 TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
|
b 670 TGGTCTCTGCCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAG 729
|
Y 141 MetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
|
b 730 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCATCTG 789
|
Y 161 AspValValGluIleThrGluAsnSerGlnGlu-LeuArgMetArgAsnSerSerGlyPr 180
|
b 790 GATGTTGTTGAAATCACAGAAACAGCCAGGAGCTTCGCATGCGGAACCAAGTGGGCC 849
|
Y 180 oCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysTh 200
|
b 850 CTGTCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTGCTGTGGAGAGCCTGAAGAC 909
|
Y 200 rProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIl 220
|
b 910 CCCCCTGTGTGGTGGGAGGAGGCCCTCTGTGATCTTGGCCCTTGGCAGGTCAGCAT 969
|
Y 220 eGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuTh 240
|
b 970 CCAGTACGACAAACAGCAGCTGTGTGGAGGAGCATCTGGACCCCTGCTCTCAC 1029
|
Y 240 rAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySe 260
|
b 1030 GGCAGCCCACTGCTTCAGGAACATACCGATGTGTCACTGGAAGTTCGGGGCAGGCTC 1089
|
Y 260 rAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAs 280
|
b 1090 AGACAAACTGGGCAGCTTCCCATCCCTGGCTGGCTGGCCAAAGATCATCATTAATTCAA 1149
|
Y 280 nProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPh 300
|
b 1150 CCCCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCCTCACTT 1209
|
Y 300 eSerGlyThrValArgProLysCysLeuProPhePheAspGluGluLeuThrProAlaTh 320
|
b 1210 CTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCAC 1269
|
Y 320 rProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIl 340
|
b 1270 CCACCTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGCTGACAT 1329
|
Y 340 eLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTy 360
|
b 1330 ACTGCTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCATTCACAGCATGCGTA 1389
|
Y 360 rGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspTh 380
|
b 1390 CCAGGGGAAGTCAACGAGAAGATGATGTGTGACAGCATCCCGAAGGGGTGTGGACAC 1449
|
Y 380 rCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGl 400
|
b 1450 CTGCCAGGGTGACAGTGGTGGGCCCTGATGTATCAATCTGACCAGTGGCATGTGGTGG 1509
|
Y 400 yIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVa 420
|
b 1510 CATCGTAGCTGGGCTATGGCTCGGGGCCCCAGCACCCAGGAGTATACACCAAGGT 1569
|
Y 420 lSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
|
b 1570 CTCAGCCTATCTCAACTGGATCTACAAATGTCTGGGAAGGCTGAGCTG 1615
|
```

```
RESULT 6
US-09-851-588-5
; Sequence 5, Application US/09851588
; Patent No., US20020042067A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND P
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DJB/JUD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2079
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-851-588-5

Alignment Scores:
Pred. No.: 3,67e-281 Length: 2079
Score: 2324.00 Matches: 434
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.23% Indels: 1
DB: 9 Gaps: 0
```

US-09-607-745-2 (1-435) x US-09-851-588-5 (1-2079)

```
Qy 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
|
Db 217 GATCCTGACAGTCATCAACCTCTGAACAGCCTCGATGTCAAACCCCTGCGCAAAACCCCGT 276
|
Qy 22 IleProMetGluThrPheArgLys-ValGlyIleProIleIleIleAlaLeuLeuSerLe 41
|
Db 277 ATCCCCATGGAGACCTTCAGAAAGTGTGGGATCCCCATCATCATAGCACTACTGAGCCT 336
|
Qy 41 uAlaSerIleIleValValValIleLysValIleLysValIleLysValIleLysValIle 61
|
Db 337 GCGAGTATCATCATTTGTTGTTCTCATCAAGGTGATTCGATATAAATACTACTTCTCT 396
|
Qy 61 uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCy 81
|
Db 397 CTGCGGGCAGCCTTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGACTG 456
|
Qy 81 sProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAl 101
|
Db 457 TCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGCCCTGCAGTGGC 516
|
Qy 101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
|
Db 517 AGTCCGCCCTCTCCAAGGACCGATCCACACTGCAGGTGCTGGACTCGGCCACAGGGAAC 576
|
Qy 121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
|
Db 577 GTTCTCTGCCCTGTTTCGACAACTTCACAGAAAGCTCTCGCTGAGACAGCCTGTAGGCAG 636
|
Qy 141 tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs 161
|
Db 637 GGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCATCTGGA 696
|
Qy 161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181
|
Db 697 TGTGTTGAAATCACAGAAACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCCCTG 756
|
Qy 181 sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr 201
|
```


b 757 TCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTGCCTGTGGGAAGACGCTGAAGACCCC 816
y 201 oArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleG1 221
b 817 CCGTGTGGTGGTGGGAGGAGGCCTCTGTGGATTCTTGGCCTTGGCAGGTCAAGATCCA 876
y 221 nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl 241
b 877 GTACGACAAACAGCACGCTCTGTGGAGGGAGCATCTGGACCCCTGAGTGGGTCTCACGGC 936
y 241 aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261
b 937 AGCCCACTGCTTCAGGAAACATACCGATGTGTCAACTGGAAGGTGCGGCGAGGCTCAGA 996
y 261 pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPr 281
b 997 CAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATCATGAATTCACCC 1056
y 281 oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301
b 1057 CATGTACCCCAAGACATGACATGCGCCCTCATGAAGCTGCAGTTCCCACTCACTTCTC 1116
y 301 rGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPr 321
b 1117 AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCC 1176
y 321 oLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLe 341
b 1177 ACTCTGGATCATTTGGATGGGGCTTTACGAAGCAGATGGAGGAAGATGTCTGACATACT 1236
y 341 uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGl 361
b 1237 GCTGCAGGCGTCAGTCCAGGTCATTCAGACACACCGTGCAATGCAGACGATGCGTACCA 1296
y 361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381
b 1297 GGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGAAGGGGGTGTGGACACCTG 1356
y 381 sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyI1 401
b 1357 CCAGGGTGACAGTGGTGGGCCCTGTATGTACCAATCTGACCAGTGGCATGTGGTGGCAT 1416
y 401 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe 421
b 1417 CGTGTAGCTGGGCTATGGCTGCGGGGCCCCGAGCACCCAGGAGTATACACCAAGGTCTC 1476
y 421 rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
b 1477 AGCCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1519

ESULT 7

S-09-776-191-71
Sequence 71, Application US/09776191
Publication No. US20030119168A1
GENERAL INFORMATION:
APPLICANT: Edwin L. Madison
APPLICANT: Edgar O. Ong
APPLICANT: Jiunn-Chern Yeh
APPLICANT: Corvas International, Inc.
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
TITLE OF INVENTION: METHODS BASED THEREON
FILE REFERENCE: 24745-1607
CURRENT APPLICATION NUMBER: US/09/776,191
CURRENT FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/213,124
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 60/234,840
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 60/179,982
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: 60/183,542
PRIOR FILING DATE: 2000-02-18

; PRIOR APPLICATION NUMBER: 09/657,968
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 2079
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (251)...(1522)
; OTHER INFORMATION: Nucleotide sequence encoding transmembrane
; OTHER INFORMATION: protease, serine 4 (TMPRSS4)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank NM016425
; DATABASE ENTRY DATE: 2000-11-06
US-09-776-191-71

Alignment Scores:
Pred. No.: 3.67e-281 Length: 2079
Score: 2324.00 Matches: 434
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.23% Indels: 1
DB: 11 Gaps: 0

US-09-607-745-2 (1-435) x US-09-776-191-71 (1-2079)
Qy 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
Db 217 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAACCCCTGCGCAACCCCGT 276
Qy 22 IleProMetGluThrPheArgLys-ValGlyIleProIleIleAlaLeuLeuSerLe 41
Db 277 ATCCCCATGGAGACCTTCAGAAAGTGTGGGATCCCCATCATCATAGCACTACTGAGCCT 336
Qy 41 uAlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLe 61
Db 337 GCGAGTATCATCATTTGGTTGTCTCTCATCAAGGTGATTCTGGATAAATACTACTTCT 396
Qy 61 uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCy 81
Db 397 CTGCGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGAGAGCTGGACTG 456
Qy 81 sProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAl 101
Db 457 TCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCCTGCAGTGGC 516
Qy 101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
Db 517 AGTCCGCTCTCCAAAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGAACTG 576
Qy 121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
Db 577 GTTCTCTGCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAGAT 636
Qy 141 tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs 161
Db 637 GGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACAGGATCTGGA 696
Qy 161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181
Db 697 TGTGTTGAAATCACAGAAAACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCCCTG 756
Qy 181 sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr 201
Db 757 TCTCTCAGGCTCCCTGGTCTCCCTGCCTGTCTTGGCTGTGGGAAGAGCCTGAAGACCCC 816
Qy 201 oArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGl 221
Db 817 CCGTGTGTGGTGGGAGGAGGCCCTCTGTGGATTCTTGGCCTTGGCAGGTCAAGCATCCA 876
Qy 221 nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl 241

b 877 GTACGACAAACAGCAGCTCTGTGGAGGGAGCATCTTGACCCCACTGGTCTCTCACGGC 936
Y 241 aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261
b 937 AGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGTGGCGGAGGCTCAGA 996
Y 261 pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPr 281
b 997 CAAACTGGGAGGCTCCCATCCCTGGCTGTGGCCAAAGATCATCATCATGAATTCACCC 1056
Y 281 oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301
b 1057 CATGTACCCCAAGACAATGACATGCGCCCTCATGAAGCTGCAGTTCACACTCATTCTC 1116
Y 301 xGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPr 321
b 1117 AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCC 1176
Y 321 oLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLe 341
b 1177 ACTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACAT 1236
Y 341 uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGl 361
b 1237 GCTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGCAATGCAGACGATGCGTAC 1296
Y 361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381
b 1297 GGGGGAAGTCACCGAAGATGATGTGTGCAGGCATCCCGAAGGGGTGTGGACACCTG 1356
Y 381 sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValGlyIle 401
b 1357 CCAGGTGACAGTGTGGCCCTTGATGTACCAATCTGACCAAGTGGCATGTGTGGGCAT 1416
Y 401 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe 421
b 1417 CGTTAGTGGGCTATGGCTGGGGGCGGAGCACCCAGGAGTATACCAAGGTCTC 1476
Y 421 rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
b 1477 AGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1519

RESULT 8

S-10-264-820-22
Sequence 22, Application US/10264820
Publication No. US20030108926A1
GENERAL INFORMATION:
APPLICANT: Eos Biotechnology, Inc.
TITLE OF INVENTION: No. US20030108926A1e1 Methods of Diagnosing Colorectal Cancer,
TITLE OF INVENTION: Compositions, and Methods of Screening for Colorectal
TITLE OF INVENTION: Cancer Modulators
FILE REFERENCE: 018501-006141US
CURRENT APPLICATION NUMBER: US/10/264,820
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: US 09/268,866
PRIOR FILING DATE: 1999-03-15
PRIOR APPLICATION NUMBER: US 09/435,945
PRIOR FILING DATE: 1999-11-09
PRIOR APPLICATION NUMBER: US 09/436,983
PRIOR FILING DATE: 1999-11-09
PRIOR APPLICATION NUMBER: US 09/450,857
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: US 09/453,850
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: US 09/493,444
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: US 09/525,993
PRIOR FILING DATE: 2000-03-15
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 22
LENGTH: 2079

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CJA8 cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (251)..(1522)
; OTHER INFORMATION: human CGA8
US-10-264-820-22
Alignment Scores:
Pred. No.: 3.67e-281 Length: 2079
Score: 2324.00 Matches: 434
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.23% Indels: 1
DB: 14 Gaps: 0
US-09-607-745-2 (1-435) x US-10-264-820-22 (1-2079)
QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
Db 217 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTCGCGCAACCCCGT 276
QY 22 IleProMetGluThrPheArgLys-ValGlyIleProIleIleIleAlaLeuLeuSerLe 41
Db 277 ATCCCCATGGAGACCTTCAGAAAGTGTGGGATCCCCATCATCATAGCACTACTGAGCCT 336
QY 41 uAlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLe 61
Db 337 GCGGAGTATCATATTGTGTTGTCTCATCAAGGTGATTCTGGATAATACTACTTCTCT 396
QY 61 uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCy 81
Db 397 CTGCGGGCAGCCTCTCCACTTCATCCGAGGAAGCAGCTGTGTACGGAGAGCTGGACTG 456
QY 81 sProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAl 101
Db 457 TCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCCTGCAGTGGC 516
QY 101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
Db 517 AGTCCGCTCTCCAAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGGAACGTG 576
QY 121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
Db 577 GTTCTCTGCTGTGTTTCGACAACTTCAGAAAGCTCTCGCTGACACAGCCTGTAGGCAGAT 636
QY 141 tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs 161
Db 637 GGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGSCCCAGACCAGGATCTGGA 696
QY 161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181
Db 697 TGTGTTGAAATCACAGAAACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCCCTG 756
QY 181 sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr 201
Db 757 TCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGCCTGTGGGAAGAGCCTGAAGACCCC 816
QY 201 oArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGl 221
Db 817 CCGTGTGTGGTGGGGAGGAGGCCCTCTGTGGATTCTTGGCCTTGGCAGGTCCAGCATCCA 876
QY 221 nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl 241
Db 877 GTACGACAAACAGCACGCTCTGTGGAGGAGCATCTCTGGACCCCACTGGGTCTCTACGGC 936
QY 241 aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261
Db 937 AGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCAGA 996
QY 261 pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPr 281

```

> 997 CAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATTTGAATTCACACCC 1056
> 281 oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301
> 1057 CATGTACCCCAAGACAATGACATGCCCTCATGAAGCTGCAGTCCCACACTACTTCTC 1116
> 301 rGlyThrValArgProIleCysLeuPropPhePheAspGluLeuThrProAlaThrPr 321
> 1117 AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCC 1176
> 321 oLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLe 341
> 1177 ACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACT 1236
> 341 uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGl 361
> 1237 GCTGCAGGCGTCAGTCCAGGTCAATGACAGCACACGGTGCAATGCAGACGATGGGTACCA 1296
> 361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381
> 1297 GGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGAAGGGGTGTGCACACCTG 1356
> 381 sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIl 401
> 1357 CCAGGTGACAGTGTGGGCCCCCTGATGTACCAATCTGCACAGTGGCATGTGTGGGCAT 1416
> 401 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe 421
> 1417 CGTTAGTGGGCTATGGCTGCGGGGGCCCCAGCACCCAGGAGTATACACCAAGGTCTC 1476
> 421 rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
> 1477 AGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1519

RESULT 9
3-10-254-289-1
Sequence 1, Application US/10254289
Publication No. US20030118509A1
GENERAL INFORMATION:
APPLICANT: Mack, David
APPLICANT: Gish, Kurt
APPLICANT: Wilson, Keith
TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
FILE REFERENCE: A-69108/DJB/JJD/AMS
CURRENT APPLICATION NUMBER: US/10/254,289
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: US/09/656,002
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: US 09/525,993
PRIOR FILING DATE: 2000-03-15
PRIOR APPLICATION NUMBER: US 09/493,444
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: PCT/US 00/07044
PRIOR FILING DATE: 2000-03-15
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 2079
TYPE: DNA
ORGANISM: Homo sapiens
3-10-254-289-1

Ligment Scores:
red. No.: 3.67e-281 Length: 2079
core: 2324.00 Matches: 434
Percent Similarity: 99.77% Conservative: 0
Percent Local Similarity: 99.77% Mismatches: 0
Very Match: 99.23% Indels: 1
3: 14 Gaps: 0

5-09-607-745-2 (1-435) x US-10-254-289-1 (1-2079)
```

```

QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
Db 217 GATCCTGACAGTGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGAAACCCCGT 276
QY 22 IleProMetGluThrPheArgLys-ValGlyIleProIleIleIleAlaLeuLeuSerLe 41
Db 277 ATCCCATGGAGACCTTCAGAAAGTGTGGGATCCCCATCATCATAGACTACTGAGCCT 336
QY 41 uAlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLe 61
Db 337 GGCAGTATCATCATTTGCTTCTCATCAAGGTGATTTCTGGATAAATACTACTTCTCT 396
QY 61 uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCy 81
Db 397 CTGCGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTACGGAGAGCTGGACTG 456
QY 81 sProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAl 101
Db 457 TCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGCCCTGCAGTGGC 516
QY 101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
Db 517 AGTCGCGCTCTCCAAGGCCGATCCACACTGCAGCTGTGGACTCGGCCACAGGAACTG 576
QY 121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
Db 577 GTTCTCTGCCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAGAT 636
QY 141 tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs 161
Db 637 GGGCTACAGCAGCAAAACCACTTTCAGAGCTGTGGAGATTGGCCCCAGACCAGGATCTGGA 696
QY 161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181
Db 697 TGTGTGTAATCACAGAAAACAGCCAGAGCTTCGCATCGGAACCTCAAGTGGGCCCTG 756
QY 181 sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr 201
Db 757 TCTCTCAGGCTCCCTGGTCTCCTGCCTGCCTGTGGAGAGAGCCTGGAAGACCC 816
QY 201 oArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGl 221
Db 817 CCGTGTGGTGGTGGGGAGGAGCCCTCTGTGGATTCTTGGCCTTGGCAGTGCAGCATCCA 876
QY 221 nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl 241
Db 877 GTACGACAAACAGCACGCTGTGTGGAGGAGCATCTTGGACCCCCACTGGTCTCTCACGGC 936
QY 241 aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261
Db 937 AGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCAGA 996
QY 261 pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPr 281
Db 997 CAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCAGAGTGCAGTTCCCACTCACTTTCTC 1056
QY 281 oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301
Db 1057 CATGTACCCCAAGACAATGACATGCCCTCATGAAGTGCAGTTCCCACTCACTTTCTC 1116
QY 301 rGlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAlaThrPr 321
Db 1117 AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCC 1176
QY 321 oLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLe 341
Db 1177 ACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACT 1236
QY 341 uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGl 361
Db 1237 GCTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGCATATGCAGACGATGGGTACCA 1296
```

361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381
1297 GGGGAAGTCAACGAGAAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACCTG 1356
381 sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
1357 CCAGGTGACAGTGGTGGCCCTGATGTACCAATCTGACCAAGTGGCATGTGTGGGCAT 1416
401 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe 421
1417 CGTTAGCTGGGCTATGGCTGCGGGGGCCGAGCACCCAGGAGTATACACCAAGGTCTC 1476
421 rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
1477 AGCCTAICTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1519

RESULT 10
S-09-776-191-3
Sequence 3, Application US/09776191
Publication No. US20030119168A1
GENERAL INFORMATION:
APPLICANT: Edwin L. Madison
APPLICANT: Edgar O. Ong
APPLICANT: Jiumn-Chern Yeh
APPLICANT: Corvas International, Inc.
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
TITLE OF INVENTION: METHODS BASED THEREON
FILE REFERENCE: 24745-1607
CURRENT APPLICATION NUMBER: US/09/776,191
CURRENT FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/213,124
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 60/234,840
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 60/179,982
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: 60/183,542
PRIOR FILING DATE: 2000-02-18
PRIOR APPLICATION NUMBER: 09/657,968
PRIOR FILING DATE: 2000-02-08
NUMBER OF SEQ ID NOS: 72
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 3
LENGTH: 2137
TYPE: DNA
ORGANISM: Homo Sapien
FEATURE:
NAME/KEY: CDS
LOCATION: (261)...(1574)
OTHER INFORMATION: DNA sequence encoding a transmembrane serine
OTHER INFORMATION: protease (MTSP3) protein
S-09-776-191-3

alignent Scores:
red. No.: 1.63e-280 Length: 2137
core: 2319.00 Matches: 432
ercent Similarity: 99.54% Conservative: 0
est Local Similarity: 99.54% Mismatches: 2
very Match: 99.02% Indels: 0
B: 11 Gaps: 0
S-09-607-745-2 (1-435) x US-09-776-191-3 (1-2137)
Y 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
b 270 GATCCTGACAGTGCATCAACCTCTGAACAGCTCGATGTCAAACCCCTGCGCAACCCCGT 329
Y 22 IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeu 41
b 330 ATCCCCATGGAGACCTTCAGAAAGGTGGGATCCCCATCATCATAGCACTACTGAGCCTG 389
Y 42 AlaSerIleIleValValValLeuIleValValLeuAspLysTyrTyrPheLeu 61

Db 390 GCGAGTATCATATTGTGGTGTCTCTCATCAAGGTGATTCTGGATAAATACTACTTCCTC 449
Qy 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
Db 450 TCGGGGCAGCCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGAGAGCTGGACTGT 509
Qy 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
Db 510 CCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGCTGCAGTGGCA 569
Qy 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
Db 570 GTCCGCTCTTCCAAAGGACCGATCCACACTGCAAGTGTGGACTCGGCCACAGGGAACCTGG 629
Qy 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
Db 630 TTCTCTGCCTGTTTCGACAACTTCACAGAAGCTCTGCTGAGACAGCCTGTAGGCAGATG 689
Qy 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
Db 690 GGTACAGCAGCAACCCACCTTCAGAGCTGTGAGATTGGCCAGACAGGATCTGGAT 749
Qy 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
Db 750 GTTGTGAAATCACAGAAACAGCCAGGAGCTTCGATCGGAACTCAAGTGGGCCCTGT 809
Qy 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
Db 810 CTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCCTGTGGAAAGAGCTGAAGACCCCC 869
Qy 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
Db 870 CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGGTCCAGCATCCAG 929
Qy 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
Db 930 TACGACATACAGCACGCTCTGTGGAGGAGCATCTCGGACCCCTTGGGCTCTCCACGGCA 989
Qy 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
Db 990 GCCCACTGCTTCAGGAAACATACCGATGTGTCTCACTGGAAGGTGCGGGCAGGCTCAGAC 1049
Qy 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
Db 1050 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATATTGAATCAACCCC 1109
Qy 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLysLeuGlnPheProLeuThrPheSer 301
Db 1110 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGTGCAGTTCCCACTACTTTCTCA 1169
Qy 302 GlyThrValArgProIleCysLysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
Db 1170 GGCACAGTCAGGCTCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1229
Qy 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
Db 1230 CTCTGGATCATTTGATGGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATACTG 1289
Qy 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 361
Db 1290 CTGCAGGGCTCAGTCCAGGTCATTGACAGCACACCGGTGCAATGCAGACGATCGGTACCAG 1349
Qy 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
Db 1350 GGGGAAGTCAACCGAGAAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACCTGC 1409
Qy 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
Db 1410 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAAGTGGCATGTGTGGGCATC 1469
Qy 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421

b 1470 GTTAGCTGGGCTATGGCTCGGGGGCCCGAGCACCCCGAGGTATACACCAAGGTCTCA 1529
y 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
b 1530 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1571
ESULT 11
S-09-888-257A-2
Sequence 2, Application US/09888257A
Publication No. US20030060612A1
GENERAL INFORMATION:
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Hillan,Kenneth J.
APPLICANT: Polakis,Paul
APPLICANT: Smith,Victoria
APPLICANT: Wood,William I.
APPLICANT: Wu,Thomas D.
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
TREATMENT OF TUMOR
FILE REFERENCE: P5002R1
CURRENT APPLICATION NUMBER: US/09/888,257A
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/063,540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: US 60/089,653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: US 60/099,792
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: US 60/103,678
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: US 60/235,451
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: PCT/US99/20111
PRIOR FILING DATE: 1999-09-01
PRIOR APPLICATION NUMBER: PCT/US00/04342
PRIOR FILING DATE: 2000-02-18
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: PCT/US00/23328
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: 2000-12-01
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: PCT/US01/06666
PRIOR FILING DATE: 2001-03-01
NUMBER OF SEQ ID NOS: 10
SEQ ID NO 2
LENGTH: 2063
TYPE: DNA
ORGANISM: Homo Sapien
S-09-888-257A-2
Lignment Scores:
red. No.: 7.76e-278 Length: 2063
core: 2297.50 Matches: 429
ercent Similarity: 98.85% Conservative: 0
est Local Similarity: 98.85% Mismatches: 0
ery Match: 98.10% Indels: 5
3: 11 Gaps: 1
3-09-607-745-2 (1-435) x US-09-888-257A-2 (1-2063)
/ 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
b 219 GATCCTGACAGTCACTCTGAAACAGCCTCGATGTCAACCCCTGCGCAACCCCGT 278

QY 22 ileProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSerLeu 41
Db 279 ATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATAGCACTACTAGCCTG 338
QY 42 AlaSerIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
Db 339 GCGAGTATCATATTGTGTGTCTCATCAAGGTGATTCTGGATAAATACTACTTCTC 398
QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
Db 399 TCGGGGAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGACTGT 458
QY 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
Db 459 CCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCCTGCAGTGGCA 518
QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
Db 519 GTCCGCCTCTCCAAGGACCGATCCACTGCAGGTGTGGACTCGGCCACAGGAACTGG 578
QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
Db 579 TTCTCTGCCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGCAGATG 638
QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
Db 639 GGCTACAGC-----AGAGCTGTGGAGATTGGCCAGACCAGGATCTGGAT 683
QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
Db 684 GTTGTGAAATCACAGAAACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCCCTGT 743
QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
Db 744 CTCTCAGGCTCCCTGGTCTCCCTGCCTGTCTTGCCTGTGGAAAGAGCCTGAAGACCCCC 803
QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
Db 804 CGTGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGGTTCAGCATCCAG 863
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
Db 864 TACGACAAACAGCACGCTCTGTGAGGGAGCATCTCTGGACCCCCACTGGGTCTCTACGGCA 923
QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
Db 924 GCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGTCAGAC 983
QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
Db 984 AAACCTGGGCAGCTTCCCAATCCCTGGCTGTGGCCAAAGATCATCATTTGAATCAACCCC 1043
QY 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
Db 1044 ATGTACCCCAAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCCCACTCACTTCTCA 1103
QY 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
Db 1104 GGCACAGTCAGGCCCATCTGTGTGCCCTTCTTTGTATGAGGAGCTCACTCCAGCCACCCCA 1163
QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
Db 1164 CTCTGGATCATTTGGATGGGGCTTTACGAAGCAGATGGAGGGAAGATGTCTGACATACTG 1223
QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 361
Db 1224 CTGCAGGCTCAGTCCAGGTCAATTGACAGCACACGGTCCAATGCAGACGATGCGTACCAG 1283
QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
Db 1284 GGGGAAGTCAACCGAGAAGATGATGTGTGCAGGCATCCCGGAAGGGGTGTGGACACCTGC 1343

Y 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
C 1344 CAGGGTGACAGTGGTGGGCCCTGATGATACCAATCTGACCACTGGCATGTGGTGGGCATC 1403
Y 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
C 1404 GTTAGCTGGGGCTATGGCTGCGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1463
Y 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
C 1464 GCCTATCTCACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505

RESULT 12

S-09-946-374-274
Sequence 274, Application US/09946374
Publication No. US20030073129A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
Acids Encoding the Same

FILE REFERENCE: P2830P1C1

CURRENT APPLICATION NUMBER: US/09/946,374

CURRENT FILING DATE: 2001-09-04

PRIOR APPLICATION NUMBER: 60/098716

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098723

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098749

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098750

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098803

PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/098821

PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/098843

PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/099536

PRIOR FILING DATE: 1998-09-09

PRIOR APPLICATION NUMBER: 60/099596

PRIOR FILING DATE: 1998-09-09

PRIOR APPLICATION NUMBER: 60/099598

PRIOR FILING DATE: 1998-09-09

PRIOR APPLICATION NUMBER: 60/099602

PRIOR FILING DATE: 1998-09-09

PRIOR APPLICATION NUMBER: 60/099642

PRIOR FILING DATE: 1998-09-09

PRIOR APPLICATION NUMBER: 60/099741

PRIOR FILING DATE: 1998-09-10

PRIOR APPLICATION NUMBER: 60/099754

PRIOR FILING DATE: 1998-09-10

PRIOR APPLICATION NUMBER: 60/099763

PRIOR FILING DATE: 1998-09-10

Y PRIOR APPLICATION NUMBER: 60/099792
C PRIOR FILING DATE: 1998-09-10
Y PRIOR APPLICATION NUMBER: 60/099808
C PRIOR FILING DATE: 1998-09-10
Y PRIOR APPLICATION NUMBER: 60/099812
C PRIOR FILING DATE: 1998-09-10
Y PRIOR APPLICATION NUMBER: 60/099815
C PRIOR FILING DATE: 1998-09-10
Y PRIOR APPLICATION NUMBER: 60/099816
C PRIOR FILING DATE: 1998-09-10
Y PRIOR APPLICATION NUMBER: 60/100385
C PRIOR FILING DATE: 1998-09-15
Y PRIOR APPLICATION NUMBER: 60/100388
C PRIOR FILING DATE: 1998-09-15
Y PRIOR APPLICATION NUMBER: 60/100390
C PRIOR FILING DATE: 1998-09-15
Y PRIOR APPLICATION NUMBER: 60/100584
C PRIOR FILING DATE: 1998-09-16
Y PRIOR APPLICATION NUMBER: 60/100627
C PRIOR FILING DATE: 1998-09-16
Y PRIOR APPLICATION NUMBER: 60/100661
C PRIOR FILING DATE: 1998-09-16
Y PRIOR APPLICATION NUMBER: 60/100662
C PRIOR FILING DATE: 1998-09-16
Y PRIOR APPLICATION NUMBER: 60/100664
C PRIOR FILING DATE: 1998-09-16
Y PRIOR APPLICATION NUMBER: 60/100683
C PRIOR FILING DATE: 1998-09-17
Y PRIOR APPLICATION NUMBER: 60/100684
C PRIOR FILING DATE: 1998-09-17
Y PRIOR APPLICATION NUMBER: 60/100710
C PRIOR FILING DATE: 1998-09-17
Y PRIOR APPLICATION NUMBER: 60/100711
C PRIOR FILING DATE: 1998-09-17
Y PRIOR APPLICATION NUMBER: 60/100848
C PRIOR FILING DATE: 1998-09-18
Y PRIOR APPLICATION NUMBER: 60/100849
C PRIOR FILING DATE: 1998-09-18
Y PRIOR APPLICATION NUMBER: 60/100919
C PRIOR FILING DATE: 1998-09-17
Y PRIOR APPLICATION NUMBER: 60/100930
C PRIOR FILING DATE: 1998-09-17
Y PRIOR APPLICATION NUMBER: 60/101014
C PRIOR FILING DATE: 1998-09-18
Y PRIOR APPLICATION NUMBER: 60/101068
C PRIOR FILING DATE: 1998-09-18
Y PRIOR APPLICATION NUMBER: 60/101071
C PRIOR FILING DATE: 1998-09-18
Y PRIOR APPLICATION NUMBER: 60/101279
C PRIOR FILING DATE: 1998-09-22
Y PRIOR APPLICATION NUMBER: 60/101471
C PRIOR FILING DATE: 1998-09-23
Y PRIOR APPLICATION NUMBER: 60/101472
C PRIOR FILING DATE: 1998-09-23
Y PRIOR APPLICATION NUMBER: 60/101474
C PRIOR FILING DATE: 1998-09-23
Y PRIOR APPLICATION NUMBER: 60/101475
C PRIOR FILING DATE: 1998-09-23
Y PRIOR APPLICATION NUMBER: 60/101476
C PRIOR FILING DATE: 1998-09-23
Y PRIOR APPLICATION NUMBER: 60/101477
C PRIOR FILING DATE: 1998-09-23
Y PRIOR APPLICATION NUMBER: 60/101479
C PRIOR FILING DATE: 1998-09-23
Y PRIOR APPLICATION NUMBER: 60/101738
C PRIOR FILING DATE: 1998-09-24
Y PRIOR APPLICATION NUMBER: 60/101741
C PRIOR FILING DATE: 1998-09-24
Y PRIOR APPLICATION NUMBER: 60/101743
C PRIOR FILING DATE: 1998-09-24
Y PRIOR APPLICATION NUMBER: 60/101915
C PRIOR FILING DATE: 1998-09-24
Y PRIOR APPLICATION NUMBER: 60/101916

PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/102207
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102240
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102307
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102330
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102331
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102484
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102487
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102570
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102571
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102684
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102687
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102965
PRIOR FILING DATE: 1998-10-02
PRIOR APPLICATION NUMBER: 60/103258
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103314
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103315
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103328
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103395
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103396
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103401
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103449
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103633
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103678
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103679
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103711
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/104257
PRIOR FILING DATE: 1998-10-14
PRIOR APPLICATION NUMBER: 60/104987
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105000
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105002
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105104
PRIOR FILING DATE: 1998-10-21
PRIOR APPLICATION NUMBER: 60/105169
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105266
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105693
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105694
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105807

Alignment Scores:

red. No.: 7,76e-278 Length: 2063
score: 2297.50 Matches: 429
percent Similarity: 98.85% Conservative: 0

Best Local Similarity: 98.85% Mismatches: 0
Query Match: 98.10% Indels: 5
DB: 11 Gaps: 1
US-09-607-745-2 (1-435) x US-09-946-374-274 (1-2063)
QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
Db 219 GATCCTGACAGTGATCAACCTGTGAACAGCCTCGATGTCAAAACCCCTGCGCAACCCCGT 278
QY 22 IleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSerLeu 41
Db 279 ATCCCGATGGAGACCTTCAGAAAGGTGGGATCCCCATCATATAGCACTACTGAGCCTG 338
QY 42 AlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
Db 339 GCGAGTATCATCATTTGGTTGTCTCATCAAGGTGATCTGGATAAATACTACTTCTC 398
QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
Db 399 TCGGGGAGCCTCTCCACTTCCATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGACTGT 458
QY 82 ProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaValAla 101
Db 459 CCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCCTGCAAGTGGCA 518
QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
Db 519 GTCCGCTCTCCAAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGGAACCTGG 578
QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
Db 579 TTCTCTGCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAGATG 638
QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
Db 639 GGCTACAGC-----AGAGCTGTGGAGATTGGCCACAGGATCTGGAT 683
QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
Db 684 GTTGTGAAATCACAGAAACAGCCAGGAGCTTCGCATGCGGAACTCAAGTGGCCCTGT 743
QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
Db 744 CTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCCTGTGGGAAGAGCCTGAAGACCCCC 803
QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
Db 804 CGTGTGGTGGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTTCAGCATCCAG 863
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
Db 864 TAGGACAAACAGCACGCTGTGTGGAGGGAGCATCTCTGGACCCCTCTGGTCTCTCAGGCA 923
QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
Db 924 GCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGAGGCTCAGAC 983
QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
Db 984 AAACCTGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTTGAATTCAACCCC 1043
QY 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
Db 1044 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGTGCAGTTCCTCCACTCACTTCTCA 1103
QY 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
Db 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163
QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
Db 1164 CTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAAATGGAGGGGAAGATGTCTGACATACTG 1223

y 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 361
b 1224 CTGAGGCGTCACTCCAGGTCAATTGACAGCACACGGTGCATGCAGACCGATCGGTACCAG 1283
y 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
b 1284 GGGGAAGTCACCGAGAAGATGATGTGCGAGGCATCCCGGAAGGGGGTGTGGACACTGC 1343
y 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
b 1344 CAGGGTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCAATCTGACCAATCTGATGGGCATC 1403
y 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
b 1404 GTTAGCTGGGGCTATGGTGGGGGCGGAGCACCCAGGAGTATACACCAAGGTCTCA 1463
y 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
b 1464 GCCTATCTCACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505

ESULT 13
S-10-015-387A-274
Sequence 274, Application US/10015387A
Publication No. US20030135034A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan I.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2830PIC54
CURRENT APPLICATION NUMBER: US/10/015,387A
CURRENT FILING DATE: 2001-12-12
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 477
SEQ ID NO 274
LENGTH: 2063
TYPE: DNA
ORGANISM: Homo sapiens
S-10-015-387A-274

Alignment Scores:
red. No.: 7,76e-278 Length: 2063
core: 2297.50 Matches: 429
percent Similarity: 98.85% Conservative: 0
est Local Similarity: 98.85% Mismatches: 0
uery Match: 98.10% Indels: 5
B: 12 Gaps: 1
S-09-607-745-2 (1-435) x US-10-015-387A-274 (1-2063)

y 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
b 219 GATCCTGACAGTGATCAACCTCTGAACAGCCCTCGATGTCAAACCCCTGGCAACCCCGT 278
y 22 IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeu 41
b 279 ATCCCCATGGAGACCTTCAGAAAGGTGGGGATCCCCCATCATCATAGCACTACTGAGCCTG 338
y 42 AlaSerIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61

Db 339 GCGAGTATCATATTGTGTGTCTCTCATCAAGGTGATTCTGGATAAAATACTACTTCCTC 398
Qy 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
Db 399 TGCGGGCAGCCTCTCCACTTCATCCGAGGAAGCAGCTGTGTACGGAGAGCTGGACTGT 458
Qy 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
Db 459 CCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCCTCAGTGGCA 518
Qy 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
Db 519 GTCCGCCTCTCCAAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGAACCTGG 578
Qy 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
Db 579 TTCTCTGCCTGTTTCGACAACTTCACAGAAGCTTCGCTGAGACAGCCTGTAGGAGATG 638
Qy 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
Db 639 GGCTACAGC-----AGAGCTGTGGAGATTGGCCACAGACCAGGATCTGGAT 683
Qy 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
Db 684 GTTGTGAAATCACAGAAACAGCCAGGAGCTTCGCATGCGGAACTCAAGTGGGCCCTGT 743
Qy 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
Db 744 CTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCCTGTGGAGAGCCTGAAGACCCCC 803
Qy 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
Db 804 CGTGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGGTTCAGCATCCAG 863
Qy 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
Db 864 TAGCACAACACAGCACGCTCTGTGGAGGAGCATCTCTGGACCCCACTGGTCTCTCAGGCA 923
Qy 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
Db 924 GCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCAGAC 983
Qy 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
Db 984 AAACCTGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATCATGAATTCACACCC 1043
Qy 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
Db 1044 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCCACTCACTTCTCA 1103
Qy 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
Db 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGTAGAGGAGCTCACTCCAGCCACCCCA 1163
Qy 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
Db 1164 CTCTGGATCATTTGGATGGGGCTTTACGAAGCAGATGAGAGGGAAGATGTCTGACATACTG 1223
Qy 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 361
Db 1224 CTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACCGTGCATATGCAGACGATGCGTACCAG 1283
Qy 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
Db 1284 GGGGAAGTCACCGAGAAGATGATGTGTGAGGCATCCCGGAAGGGGTGTGGACACCTGC 1343
Qy 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
Db 1344 CAGGGTGACAGTGGTGGGCCCTTGATGTACCAATCTGACCAAGTGGCATGTGTGGGGCATC 1403
Qy 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
Db 1404 GTTAGCTGGGGCTATGGCTGCGGGGGCCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1463

Y 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
| | | | |
b 1464 GCCTATCTCACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1505

RESULT 14
S-10-063-735-111
Sequence 111, Application US/10063735
Publication No. US2003013882A1
GENERAL INFORMATION:
APPLICANT: Eaton, Dan L.
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3230R1C1
CURRENT APPLICATION NUMBER: US/10/063,735
CURRENT FILING DATE: 2002-05-08
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 170
SEQ ID NO 111
LENGTH: 2063
TYPE: DNA
ORGANISM: Homo Sapien
S-10-063-735-111

Alignment Scores:
red. No.: 7.76e-278 Length: 2063
core: 2297.50 Matches: 429
Percent Similarity: 98.85% Conservative: 0
est Local Similarity: 98.85% Mismatches: 0
Query Match: 98.10% Indels: 5
B: 12 Gaps: 1
S-09-607-745-2 (1-435) x US-10-063-735-111 (1-2063)

Y 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
| | | | |
b 219 GATCCTGACAGTGATCAACCTCTGAACAGCTCGATGTCAAACCCCTGGCAACCCCGT 278
| | | | |
Y 22 IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuSerLeu 41
| | | | |
b 279 ATCCCCATGGAGACCTTCAGAAAGGTGGGATCCCCATCATCATAGCACTACTGAGCCTG 338
| | | | |
Y 42 AlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
| | | | |
b 339 GCGAGTATCATATGTGTGTTGCTCTCATCAAGGTGATTTCTGGATAAATACTATTCTCTC 398
| | | | |
Y 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
| | | | |
b 399 TCGGGGAGCCCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGGAGAGCTGGACTGT 458
| | | | |
Y 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
| | | | |
b 459 CCCTTGGGGAGGAGGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTGCAGTGGCA 518
| | | | |
Y 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
| | | | |
b 519 GTCCGCCTCTCCAGGACCGATCCACACTGAGGTGTGGACTCGGCCACAGGGAACCTG 578
| | | | |
Y 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
| | | | |
b 579 TTCTCTGCCTGTTTCGACAACTTCACAGAACTCTCGCTGAGACAGCCTGTAGGAGATG 638
| | | | |
Y 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
| | | | |
b 639 GGCTACAGC-----AGAGCTGTGGAGATTGGCCCAAGACCAAGGATCTGGAT 683

QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
| | | | |
Db 684 GTTGTGTAATCACAGAAAACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCCCTGT 743
| | | | |
QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
| | | | |
Db 744 CTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCCCTGTGGAAAGAGCCCTGAAGACCCCC 803
| | | | |
QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
| | | | |
Db 804 CGTGTGGTGGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTTCAGCATCCAG 863
| | | | |
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
| | | | |
Db 864 TACGACAAACAGCACGCTCTGTGGAGGGAGCATCCTGGACCCCACTGGGTCTCTACGGCA 923
| | | | |
QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
| | | | |
Db 924 GCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCAGAC 983
| | | | |
QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
| | | | |
Db 984 AAACCTGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATCATTTGAATTCACCCC 1043
| | | | |
QY 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
| | | | |
Db 1044 ATGTACCCCAAAGACAATGACATCGCCCTCATGAAGCTGAGGTTCCCACTCACTTTCTCA 1103
| | | | |
QY 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
| | | | |
Db 1104 GGCACAGTCAGGCCCATCTGTCTGCCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163
| | | | |
QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu 341
| | | | |
Db 1164 CTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACTG 1223
| | | | |
QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 361
| | | | |
Db 1224 CTGCAGGCGTCAGTCCAGGTCTATGACAGCACACCGGTGCAATGCAGACGATCGGTACCAG 1283
| | | | |
QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
| | | | |
Db 1284 GGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGAAGGGGTGTGGACACTGC 1343
| | | | |
QY 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
| | | | |
Db 1344 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGCATC 1403
| | | | |
QY 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
| | | | |
Db 1404 GTTAGCTGGGGCTATGGCTCGGGGGCCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1463
| | | | |
QY 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
| | | | |
Db 1464 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505
| | | | |

RESULT 15
US-10-006-130A-274
; Sequence 274, Application US/10006130A
; Publication No. US20030148375A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2830P1C7
CURRENT APPLICATION NUMBER: US/10/006,130A
CURRENT FILING DATE: 2002-03-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 477
SEQ ID NO 274
LENGTH: 2063
TYPE: DNA
ORGANISM: Homo sapiens
S-10-006-130A-274
alignent Scores:
red. No.: 7.76e-278 Length: 2063
core: 2297.50 Matches: 429
ercent Similarity: 98.85% Conservative: 0
est Local Similarity: 98.85% Mismatches: 0
uery Match: 98.10% Indels: 5
B: 12 Gaps: 1
S-09-607-745-2 (1-435) x US-10-006-130A-274 (1-2063)
Y 2 AspProAspSerAspGlnProLeuAsnSerLeuAspVallylsProLeuArgLysProArg 21
b 219 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAACCCCTGCGCAACCCCGT 278
Y 22 IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeu 41
b 279 ATCCCATGGAGACCTTCAGAAAGGTGGGATCCCCATCATATAGCACTACTGAGCCTG 338
Y 42 AlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
b 339 GCGAGTATCATATGTGTTGTCTCATCAAGGTGATTCGTGGATAAATACTACTTCCTC 398
Y 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
b 399 TCGGGGAGCCTCTCCACTTCATCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGACTGT 458
Y 82 ProLeuGlyGluAspGluHisCysVallylsSerPheProGluGlyProAlaValAla 101
b 459 CCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCCTGCAGTGGCA 518
Y 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
b 519 GTCCGCCTCTCCAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGGAACTGG 578
Y 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
b 579 TTCTCTGCCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAGATG 638
Y 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
b 639 GGCTACAGC-----AGAGCTGTGGAGATTGGCCACAGCCAGGATCTGGAT 683
Y 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
b 684 GTTGTGAAATCACAGAAACAGCCAGGAGCTTCGCATCGGGAACCTCAAGTGGGCCCTGT 743
Y 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
b 744 CTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTGCCTGTGGGAAGAGCCTGAAGACCCCC 803
Y 202 ArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
b 804 CGTGTGGTGGGTGGGAGGAGGCCTCTGTGGATTCTTGGCCCTGGCAGGTCAGCATCCAG 863
Y 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
b 864 TACGACAAACAGCACGTCGTGTGGAGGAGGATCCTGGACCCCACTGGGTCTCTCAGGCA 923

QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
Db 924 GCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGCTCAGAC 983
QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
Db 984 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAATCATCATCATGAATTCAACCCC 1043
QY 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
Db 1044 ATGTACCCCAAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCCTCACTTCTCTCA 1103
QY 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
Db 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163
QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
Db 1164 CTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACTG 1223
QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 361
Db 1224 CTGCGGGCGTCAGTCCAGGTCATTGACAGCACACCGTGCAATGCAGACGATCGGTACCAG 1283
QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
Db 1284 GGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGAAGGGGTGTGGACACCTGC 1343
QY 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
Db 1344 CAGGTGACAGTGGTGGGCCCTGTATGTACCAATCTGACCAGTGCCATGTGGTGGGCATC 1403
QY 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
Db 1404 GTTAGCTGGGCTATGGCTGCGGGGGCCCGAGCACCCAGAGTATACACCAAGTCTCTCA 1463
QY 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
Db 1464 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505

Search completed: November 23, 2003, 12:22:18
Job time : 514.909 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

1 protein - nucleic search, using frame_plus_p2n model

in on: November 23, 2003, 07:53:00 ; Search time 3092.27 Seconds
(without alignments)
3418.993 Million cell updates/sec

File: US-09-607-745-2
Effect score: 2342
Sequence: 1 MDPDSQPLNSLDVKPLRKP.....VYTKVSAYLNWIYNVWKAEL 435

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Archived: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
MODEL=frame+p2n.model -DEV=xlh
J=/cgn2_1/USPTO_spool/US09607745/runat_21112003_144346_22229/app_query.fasta_1.1038
JB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
INITs=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
LOCALALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US09607745@cgn_1_1_3596 @runat_21112003_144346_22229 -NCPU=6 -ICPU=3
JO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2044	87.3	1973	11	BC004855	BC004855 Homo sapi
2	1668.5	71.2	2177	11	AK078890	AK078890 Mus muscu
3	1233	52.6	844	10	BE531100	BE531100 601278466
4	1203	51.4	895	13	BU522841	BU522841 AGENCOURT
5	1158.5	49.5	975	10	BG288427	BG288427 602388091
6	1108.5	47.3	920	10	BE616663	BE616663 601278696
7	1054.5	45.0	617	12	BM795149	BM795149 K-EST0076
8	1042	44.5	956	13	BU523218	BU523218 AGENCOURT
9	971	41.5	569	12	BM686292	BM686292 UI-E-CRO-
10	951.5	40.6	804	12	BG822914	BG822914 602727917
11	949.5	40.5	1007	10	BG386903	BG386903 602454702
12	893	38.1	920	13	BQ922656	BQ922656 AGENCOURT
13	891	38.0	689	12	BG966811	BG966811 602834306
14	890.5	38.0	879	10	BE616186	BE616186 601278758
15	885	37.8	515	14	CB142902	CB142902 K-EST0196
16	881.5	37.6	645	10	BF118991	BF118991 601755351
17	877	37.4	503	10	BE514663	BE514663 601317126
18	875.5	37.4	787	10	BE615750	BE615750 601279885
19	858	36.6	498	12	BM795128	BM795128 K-EST0076
20	856	36.5	548	14	CB159713	CB159713 K-EST0219
21	848	36.2	744	12	BM008802	BM008802 603618532
22	844	36.0	742	10	BG468475	BG468475 602510448
23	832.5	35.5	693	10	BG481239	BG481239 602528566
24	825.5	35.2	935	10	BE615106	BE615106 601280475
25	816	34.8	659	9	AI924527	AI924527 wn61c07.x
26	808	34.5	648	9	AI924182	AI924182 wn53f11.x
27	808	34.5	931	13	BU157248	BU157248 AGENCOURT
28	797	34.0	563	12	BM773306	BM773306 K-EST0057
29	797	34.0	572	12	BM746729	BM746729 K-EST0021
30	797	34.0	578	12	BM747250	BM747250 K-EST0021
31	797	34.0	593	12	BM788452	BM788452 K-EST0067
32	797	34.0	596	12	BM772743	BM772743 K-EST0056
33	797	34.0	614	12	BM788163	BM788163 K-EST0067
34	797	34.0	633	12	BM764659	BM764659 K-EST0046
35	797	34.0	691	12	BM763697	BM763697 K-EST0045
36	797	34.0	692	12	BM772754	BM772754 K-EST0056
37	797	34.0	700	12	BM772620	BM772620 K-EST0056
38	783	33.4	548	12	BM746721	BM746721 K-EST0021
39	775.5	33.1	977	10	BE614660	BE614660 601281614
40	770	32.9	462	12	BG984172	BG984172 IL5-CN006
41	765.5	32.7	1088	10	BE616293	BE616293 601281476
42	746.5	31.9	1835	11	BC048135	BC048135 Danio rer
43	740.5	31.6	1105	10	BE736425	BE736425 601307316
44	737	31.5	850	12	BI251465	BI251465 602994002
45	728	31.1	803	14	CA311909	CA311909 UI-CF-FNO

ALIGNMENTS

RESULT 1
BC004855
LOCUS
DEFINITION BC004855 1973 bp mRNA linear HTC 12-JUL-2001
Homo sapiens, Similar to transmembrane protease, serine 4, clone
IMAGE:3835263, mRNA.
ACCESSION BC004855
VERSION BC004855.1 GI:14709533
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1973)

AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: DCRD/DRP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 14 Row: p Column: 19
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8347148
This clone has the following problem: frame shifted.

FEATURES
Location/Qualifiers
source 1..1973
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3835263"
/tissue_type="Ovary, adenocarcinoma"
/clone_lib="NIH MGC 9"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

ASE COUNT 471 a 573 c 531 g 398 t
RIGIN

Lignment Scores:

red. No.: 2,51e-210 Length: 1973
score: 2044.00 Matches: 387
percent Similarity: 88.76% Conservative: 0
best Local Similarity: 88.76% Mismatches: 0
every Match: 87.28% Indels: 49
B: 11 Gaps: 1

S-09-607-745-2 (1-435) x BC004855 (1-1973)

y 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
o 243 ATGGATCCTGCAGTGATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAAAACCC 302
y 21 ArgileProMetGluThrPheArgLysValGlyIleProilellelelelelelelelelele 40
o 303 CGTATCCCCATGGAGACCTTCAGAAAGGTGGGATCCCCATCATCATAGCACTACTGAGC 362
y 41 LeuAlaSerille 60
o 363 CTGGCGAGTATCATCTGTGGTTGTCTCATCAAGGTGATTCTGGATAAATACTACTTC 422
y 61 LeuCysGlyGlnProLeuHisPheilleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
b 423 CTCTCGGGGCAGCCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGGAGAGCTGGAC 482
y 81 CysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaVal 100
b 483 TGTCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCCTGCAGTG 542
y 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
b 543 GCAGTCCGCGCTCTCCAAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGGAAC 602
y 121 TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140

Db 603 TGGTTCTCTGCCTGTTTCGACAACTTCAAGAAAGCTCTCGCTGAGACAGCCTGTAGGCAG 662
Qy 141 MetGlyTyrSerSerLysProThrPheArgAlaValGluilleGlyProAspGlnAspLeu 160
Db 663 ATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCAGACCAAGGATCTG 722
Qy 161 AspValValGluilleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180
Db 723 GATGTTGTTGAAATCACAGAAACAGCCAGAGCTTCGCATCGGAACTCAAGTGGGCC 782
Qy 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
Db 783 TGTCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCCTGTGGGAAGAGCCTGAAGACC 842
Qy 201 ProArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
Db 843 CCGCGTGTGGTGGGAGGAGGCGCTCTGTGGATTCTTGGCCTTGGCAGGTCAGCATC 902
Qy 221 GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
Db 903 CAGTACGACAAACAGCACGCTGTGTGGAGGAGCATCCTGGACCCCCACTGGGTCTCTCAG 962
Qy 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260
Db 963 GCAGCCCACTGTCTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCA 1022
Qy 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsn 280
Db 1023 GACAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCTGAGCAAGATCATCATGAATTGAATCAAC 1082
Qy 281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
Db 1083 CCCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTCCCACCTCACTTTC 1142
Qy 301 SerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThr 320
Db 1143 TCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACC 1202
Qy 321 ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle 340
Db 1203 CCACTCTGGATCATTGGATGGGGCTTTACGAAGCAGAAATGGA----- 1244
Qy 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyr 360
Db 1244 ----- 1244
Qy 361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThr 380
Db 1244 ----- 1244
Qy 381 CysGlnGly-AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValG1 400
Db 1245 -----GGGTGACAGTGGTGGCCCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGG 1298
Qy 400 YIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTrpThrLysVa 420
Db 1299 CATCGTTAGCTGGGGCTATGCTGCGGGGGCCCCGACACCCAGGAGTATACACCAAGGT 1358
Qy 420 lSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
Db 1359 CTCAGCCTATCTCAACTGGATGTACAATGTCTGGAAGGCTGAGCTG 1404
RESULT 2
AK078890 2177 bp mRNA linear HTC 05-DEC-2002
LOCUS Mus musculus adult male colon cDNA, RIKEN full-length enriched library, clone:9030622G02 product:SIMILAR TO TRANSMEMBRANE
DEFINITION PROTEASE, SERINE 4 homolog [Mus musculus], full insert sequence.
ACCESSION AK078890
VERSION AK078890.1 GI:26098158
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851

5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2177)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Y 120 nTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgG1 140
b 559 CTGGGCTCAGTCTGTTTCGACAACTTCACAGAAGCACTGGCAAGACAGCCTGCAGACA 618
Y 140 nMetGlyTyrSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLe 160
b 619 GATGGGCTATACAGCCAGCCCGCTTTCAGAGCAGTGGAGATCCGTCCAGATCAGAACCT 678
Y 160 uAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPr 180
b 679 CCCTGTTGCTCAAGTCACAGGAACACAGCCAGGAACCTTCAGGTGCAGATGGAAGCAGATC 738
Y 180 oCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysTh 200
b 739 CTGCCTCTCAGGCTCCCTGGTTTCCCTTGCCTGCCTTGACTGTGGAAGAGAGCCTGAAGAC 798
Y 200 rProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerI1 220
b 799 TCCTCGTGTGGTGGGTGGGTGGAGGCCCTGTGGATTCTTGGCCGTGGCAGGTCAAGAT 858
Y 220 eGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuTh 240
b 859 CCAGTACAACAAGCAGCATGTCTGTGGTGGGAGCATCCTGGATCCCACTGGATCCTCAC 918
Y 240 rAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySe 260
b 919 AGCAGCCCACTGCTTCAGGAAGTATCTGTATGTGTCAAGCTGGAAGGTCAAGGCGAGCTC 978
Y 260 rAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAs 280
b 979 AAACATACTGGTAACCTCTCCATCCTTGCCTGTGGCCAAAGATCTTCATCGTCGTAACCAA 1038
Y 280 nProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPh 300
b 1039 TCCTCTGTACCCCAAGAGAAGAGACATTCCTTGTTAAGCTGCAGATGCCACTCAATT 1098
Y 300 eSerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaTh 320
b 1099 CTCAGGCTCAGTCAGGCCCATCTGCCCTGCTTCTGTATAGGTGCTTGTCCAGCCAC 1158
Y 320 rProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspI1 340
b 1159 ACCAGTCTGGTCACTGGATGGGCTTTACAGAAGAAACCGGAGAAAGATGCTGACAT 1218
Y 340 eLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTy 360
b 1219 GCTACTGCAGGCATCAGTCCAGGTCAATTGACAGCACACGGTGTGAATGCAGAGATGCTA 1278
Y 360 rGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspTh 380
b 1279 CGAAGGGGAAGTGACCGCTGAGATGCTGTGTGCAGGTACCCCAAGGGTGGCAAGGACAC 1338
Y 380 rCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValG1 400
b 1339 CTGCCAGGGTGACAGTGGTGGGCCCTTTGATGTACCATTTCTGACAAGTGGCAGGTAGG 1398
Y 400 yIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVa 420
b 1399 CATCGTAGCTGGGGCCATGGATCGGGGCCCAAGTACTCTCTGGAGTGTATATACCAAGGT 1458
Y 420 lSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
b 1459 CACTGCCTATCTCAACTGGATCTACAATGTTCGGAAGTCTGAGATG 1504

ESULT 3
E531100
OCUS BE531100 844 bp mRNA linear EST 09-AUG-2000
EFINITION 601278466F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3610337 5',
mRNA sequence.
CCESION BE531100
ERSION BE531100.1 GI:9759745
EYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 844)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM266 row: d column: 18
High quality sequence stop: 712.

FEATURES
Location/Qualifiers
1..844
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3610337"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_39"
/note="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies)."

BASE COUNT 165 a 251 c 244 g 184 t
ORIGIN
Alignment Scores:
Pred. No.: 6.12e-123 Length: 844
Score: 1233.00 Matches: 254
Percent Similarity: 91.49% Conservative: 4
Best Local Similarity: 90.07% Mismatches: 14
Query Match: 52.65% Indels: 15
DB: 10 Gaps: 2
US-09-607-745-2 (1-435) x BE531100 (1-844)

QY 40 SerLeuAlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrTyr 59
Db 1 AGCCTGGCGAGTATCATCATCTGTGGTTGTCTCTCATCAAGGTGATTCTGGATAAATACTAC 60
QY 60 PheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeu 79
Db 61 TTCCTCTGCGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGAGAGCTG 120
QY 80 AspCysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAla 99
Db 121 GACTGTCCCTTGGGGAGGAGGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCCTGCA 180
QY 100 ValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGly 119
Db 181 GTGGCAGTCCGCTCTCCAAAGACCGATCCACACTGCAGGTGTGTGACTCGCCACAGGG 240
QY 120 AsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArg 139
Db 241 AACTGGTTCTCTGCCTGTTCGACAACCTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGG 300
QY 140 GlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAsp 159
Db 301 CAGATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACAGCAGGAT 360
QY 160 LeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGly 179

361	CTGGATGTTGTTGAATCAAGAAACAGCCAGGAGCTTCGATGCGGAACCTCAAGTGGG	420
180	ProCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLys	199
421	CCCTGTCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCTGTGGGAAGAGCCTGAAG	480
200	ThrProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSer	219
481	ACCCCCCGTGTGGTGGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGGTGAGC	540
220	IleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu	239
541	ATCCAGTACGACAAACAGCACGTCTGTGGAGGAGCATCTCTGGACCCCACTGGGTCTCTC	600
240	ThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGly	259
601	ACGGCAGC-CACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGC	659
260	SerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPhe	279
660	TCAGACAA-CTGGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATTTGAATCC	718
280	AsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhePro-----	297
719	AACCTGTT--CCCCAAGACAT-GACATGCCCTCATGAAGCTGCG-TTCCCCTCACTT	773
298	-----LeuThrPheSerGlyThrValArgProIleCysLeuPro	310
774	CTCCGGCCCGCCGCCCTCTGGTTGCTCTTTGTGGGGGCCCT-CGCCCTCTGTCTGGGG	832
311	PhePhe 312	
833	GGGTTC 838	
RESULT 4		
J522841		
LOCUS	BU522841	895 bp mRNA linear EST 13-SEP-2002
DEFINITION	AGENCOURT 10158166 NCI CGAP Co24 Mus musculus cDNA clone	
	IMAGE:6529422 5', mRNA sequence.	
ACCESSION	BU522841	
VERSION	BU522841.1 GI:22833279	
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE	1 (bases 1 to 895)	
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .	
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: The Cepko Laboratory cDNA Library Preparation: Life Technologies, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LILNL at: http://image.llnl.gov Plate: LLAM14129 row: i column: 06 High quality sequence stop: 665. Location/Qualifiers	
FEATURES	1..895	
source	/organism="Mus musculus"	
	/mol_type="mRNA"	
	/strain="FVB/N"	
	/db_xref="taxon:10090"	
	/clone="IMAGE:6529422"	
	/lab_host="DH10B (T1 phage-resistant)"	
	/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.	

Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 202 a 242 C 257 G 194 t
ORIGIN

Alignment Scores:
Pred. No.: 1.2e-119 Length: 895
Score: 1203.00 Matches: 217
Percent Similarity: 89.17% Conservative: 30
Best Local Similarity: 78.34% Mismatches: 29
Query Match: 51.37% Indels: 1
DB: 13 Gaps: 0

US-09-607-745-2 (1-435) x BU522841 (1-895)

QY	154	IleGlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArg	173
Db	23	GTCCGTCAGATCAGAACCCTCCCTGTTGCTCAAGTCACAGGAAACAGCCAGGAACCTTCAG	82
QY	174	MetArgAsnSerSerGlyProCysLeuSerGlySerLeuValSerLeuHisCysLeuAla	193
Db	83	GTGCAGAATGGAAGCAGATCCTGCCTCTCAGGCTCCCTGGTTTCCTTGGCTGCCTTGAC	142
QY	194	CysGlyLysSerLeuLysThrProArgValValGlyGlyGluAlaSerValAspSer	213
Db	143	TGTGGAAGAGCCTGAAGACTCTCTGTTGTTGGTGGGTGGAGGCCCTGTGGATTCT	202
QY	214	TrpProTrpGlnValSerIleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeu	233
Db	203	TGGCCGTGGCAGGTCAAGCATCCAGTACAACAAGCAGCATGTCTGTGGTGGAGCATCCTG	262
QY	234	AspProHisTrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsn	253
Db	263	GATCCCACTGGATCCTCACAGCAGCCCACTGCTTCAGGAAGATATCTTGATGTGTCAAGC	322
QY	254	TrpLysValArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLys	273
Db	323	TGGAAGGTCAGGCGAGGCTCAACATATACTGGGTAACTCTCCATCCTTGCTGTGGCCAA	382
QY	274	IleIleIleIleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLys	293
Db	383	ATCTTCATCGCTGAACCCCAATCTCTGTACCCCAAGAGAAGACATTGCCCTTTGTTAAG	442
QY	294	LeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProphePheAsp	313
Db	443	CTGCAGATGCCACTCACATTCTCAGGCTCAGTCAGGCCCATCTGCCTTCTCTGAT	502
QY	314	GluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsn	333
Db	503	GAGGTGCTTGTCACGCCACACCAGTCTGGGTCACTTGGATGGGGCTTTACAGAAGAAAC	562
QY	334	GlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArg	353
Db	563	GGAGGAAAGATGTCTGACATGTCTGACAGGCATCAGTCCAGGTCACTCGACAGCACACGG	622
QY	354	CysAsnAlaAspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIle	373
Db	623	TGCAATGCAGAGGATGCCTACGAAGGGGAAGTGACCGCTGAGATGCTGTGTGACAGGTACC	682
QY	374	ProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSer	393
Db	683	CCACAGGTTGGCAAGGACACCTGCCAGGTTGACAGTGTGGGCCCTTTGATGTACCAATCT	742
QY	394	AspGlnTrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThr	413
Db	743	GACAGTGGCAGGTAGTAGGCATCGTGAGCTGGGGGCCATGGATCGGGCGGCCCAAGTACT	802
QY	414	-ProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsn	429
Db	803	CCCTGGAGTGTATACCAAGGTCACTGCCTATCTCAACTGGATCTACAAT	851

RESULT 5
BG288427


```

JCUS      BG288427      975 bp      mRNA      linear      EST 21-FEB-2001
DEFINITION 602388091F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4516990 5',
            mRNA sequence.
JCSESSION  BG288427
JSESSION   BG288427.1 GI:13043459
JWORDS     EST.
JURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 975)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-re@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10409 row: e column: 23
            High quality sequence stop: 659.
            Location/Qualifiers
              1..975
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4516990"
                /tissue_type="transitional cell papilloma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_93"
                /note="Organ: bladder; Vector: pCMV-SPORT6; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 1.7 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."
BASE COUNT 208 a 270 c 300 g 197 t
FEATURES
    source
        1..975
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:4516990"
            /tissue_type="transitional cell papilloma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_93"
            /note="Organ: bladder; Vector: pCMV-SPORT6; Site 1: NotI;
            Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
            Average insert size 1.7 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC Library."
BASE COUNT 208 a 270 c 300 g 197 t

```

```

Db      303 TTCCCACTCACTTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAG 362
Qy      316 LeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGly 335
Db      363 CTCACCTCCAGCCACCCCACTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAATGGAGG 422
Qy      336 LysMetSerAspIleLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsn 355
Db      423 AAGATGTCTGACATACTGCTGCAGGCGTCACTCCAGGTCAATTGACAGCACACGGTGTCAAT 482
Qy      356 AlaAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGlu 375
Db      483 GCAGACGATCGGTACAGGGGGAAGTCCCGAGAAGATGATGTGTGCAGGCATCCCGGAA 542
Qy      376 -GlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspG1 395
Db      543 GGGGGGTGTGGACACCTGCCA-GGTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCA 601
Qy      395 nTrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProG1 415
Db      602 GTGGCATGTGGTGGGCATCGTTAGCTGGGGCTATGGCTCGGGGGCC---GAGCACCCAGG 658
Qy      415 yValTyr-ThrLysValSerAla 422
Db      659 AGTTTACCACCAAGGTCTCAAGC 681
RESULT 6
BE616663
LOCUS     BE616663      920 bp      mRNA      linear      EST 24-AUG-2000
DEFINITION 601278696F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3610990 5',
            mRNA sequence.
ACCESSION BE616663
VERSION    BE616663.1 GI:9898262
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
            Homo sapiens
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 920)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-re@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM267 row: o column: 23
            High quality sequence start: 5
            High quality sequence stop: 819.
            Location/Qualifiers
              1..920
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:3610990"
                /tissue_type="adenocarcinoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_39"
                /note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;
                Site 2: EcoRI; cDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Library constructed
                by Ling Hong in the laboratory of Gerald M. Rubin
                (University of California, Berkeley) using ZAP-cDNA
                synthesis kit (Stratagene) and Superscript II RT (Life
                Technologies)."
BASE COUNT 205 a 256 c 256 g 203 t

```



```

c 2 GACGGAGAGCTGGACTGTCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTCCCC 61
Y 96 GluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAsp 115
c 62 GAAGGCGCTGCAGTGGCAGTCCGCTCTCCAAGGACCGATCCACACTGCAGGTGCTGGAC 121
Y 116 SerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGlu 135
c 122 TCGGCCACAGGAACTGGTTCTCTGCCTGTTTCGACAACTTCACAGAAAGCTTCGCTGAG 181
Y 136 ThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGly 155
c 182 ACAGCCTGTAGGCAGATGGGCTACAGCAGCAAAACCCACTTCAGAGCTGTGGAGATTGGC 241
Y 156 ProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArg 175
b 242 CCAGACCAGGATCTGGATGTTGTTGAAATCACAGAAACAGCCAGGAGCTTCGCATGCGG 301
Y 176 AsnSerSerGlyProCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGly 195
b 302 AACTCAAGTGGGCCCTGTCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGCTGTGGG 361
Y 196 LysSerLeuLysThrProArgValValGly--GlyGluGluAlaSerValAspSerTrp 214
b 362 AAGAGCCTGAAGACCCCCCGTGTGTGGNGTGGGGAAGGANGCCCTCTGTGGATTCTTGG 421
Y 215 ProTrpGlnValSerIleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAsp 234
b 422 CCTTGGCAGGTCAGCATACAGTACGACAAACAGCACGTCGTGTGGAGGGAGCATCTGGAC 481
Y 235 ProHisTrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrp 254
b 482 CCCCACTGGGTCTCAGGCAGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGG 541
Y 255 LysValArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIle 274
b 542 AAGGTGCGGCGAGGCTCAAAACAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATC 601
Y 275 IleIleIleGluPhe 279
b 602 ATCATCATTTGAATTC 616

RESULT 8
U523218
OCUS AGENCOURT 10154487 NCI_CGAP_Co24 Mus musculus cDNA clone
EFINITION IMAGE:6529864 5', mRNA_sequence.
CCESION BU523218
ERSION BU523218.1 GI:22833656
EYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 956)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAMi4130 row: k column: 16
High quality sequence stop: 706.
Location/Qualifiers
FEATURES
source
1..956
/organism="Mus musculus"
/mol_type="mRNA"
```

```

/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6529864"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Co24"
/note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 214 a 275 c 280 g 187 t
ORIGIN
Alignment Scores:
Pred. No.: 3.87e-102 Length: 956
Score: 1042.00 Matches: 213
Percent Similarity: 81.25% Conservative: 34
Best Local Similarity: 70.07% Mismatches: 48
Query Match: 44.49% Indels: 9
DB: 13 Gaps: 4
US-09-607-745-2 (1-435) x BU523218 (1-956)
QY 96 GluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAsp 115
Db 46 GAAAGCCCGGAGTGGCAGTCCGGCTCTCCAGGACAGATCCACCCCTGCAGGTGCTGGAT 105
QY 116 SerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGlu 135
Db 106 GCAGCCACAGGGACCTGGCCCTCAGTCTGTTTCGACAACTTCACAGAAAGCACTGGCCAAG 165
QY 136 ThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGly 155
Db 166 ACAGCCTGCAGACAGATGGGCTATGACAGCCAGCCCGCTTTCAGAGCAGTGGAGATCCGT 225
QY 156 ProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArg 175
Db 226 CCAGATCAGAACCTCCCTGCTCTCAGGCTCCCTGGTTTCCCTGGCTGCCTGACTGTGGA 285
QY 176 AsnSerSerGlyProCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGly 195
Db 286 AATGGAAGCAGATCCTGCTCTCAGGCTCCCTGGTTTCCCTGGCTGCCTGACTGTGGA 345
QY 196 LysSerLeuLysThrProArgValValGlyGlyGluGluAlaSerValAspSerTrpPro 215
Db 346 AAGAGCCTGAAGACTCCTCGTGTGTGAGTGGGTGGAGGCCCTGTGGATTCTTGGCCG 405
QY 216 TrpGlnValSerIleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspPro 235
Db 406 TGGCAGGTCAGCATCCAGTACACAAAGCAGCATGTCTGTGGTGGGAGCATCTGGATCCC 465
QY 236 HisTrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLys 255
Db 466 CACTGGATCCTCACAGCAGCCCACTGCTTCAGGAAGATATCTTGATGTGTCAAGCTGGAAG 525
QY 256 ValArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIle 275
Db 526 GTCAGGCGAGGCTCAACATATACTGGGTAACTCTCCATCCTTGCCTGTGGCCAAAGATCTTC 585
QY 276 IleIleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGln 295
Db 586 ATCGGTGAACCCCAATCCTCTGTACCCCAAGAGAGGACATTGCCCCCTTTTAAGCTGCAG 645
QY 296 PheProLeuThrPheSerGlyThrValArgProIleCysLeuProPheAspGluGlu 315
Db 646 ATGCCACTCACATTCTCAGGCTCAGTCAGGCCCATCTGCCTGCCCTTCTCTGATGAGGTG 705
QY 316 LeuThrProAlaThrProLeuTrpIleIleGlyTrp-GlyPheThrLysGlnAsn-GlyG 335
Db 706 CTTGTCCAGCCACACACAGTCTGGGTCAATGGATGGGGGCTTTACAGAAAGAAACCGGAG 765
QY 335 LysMetSerAspIleLeuGln-AlaSerVal-GlnValIleAspSerThrArgCy 354
Db 766 GAAGATGTCCTGACATGTACTGCAAGGCATCACTCCAGGTCATCGAAGCACACCGGTG 825
```


DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLCM1735 row: b column: 05
High quality sequence stop: 793.
Location/Qualifiers

FEATURES

source
1. .804
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4867108"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_15"
/note="Organ: colon; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
ASE COUNT 177 a 235 c 235 g 157 t
RIGIN
Alignment Scores:
red. No.: 1.93e-92 Length: 804
core: 951.50 Matches: 193
Percent Similarity: 97.98% Conservative: 1
Best Local Similarity: 97.47% Mismatches: 3
Query Match: 40.63% Indels: 3
B: 12 Gaps: 0
3-09-607-745-2 (1-435) x BG822914 (1-804)

y 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
o 215 ATGGATCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAACCCCTGCGCAACCC 274
y 21 ArgileProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSer 40
o 275 CGTATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 334
y 41 LeuAlaSerIleIleValValValIleLysValIleLeuAspLysTyrTyrPhe 60
b 335 CTGGCGAGTATCATATTGTGTTGTCTCATCAAGGTGATCTGGATAAATACTACTTC 394
y 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
b 395 CTCTCGGGGAGCCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGGAGAGCTGGAC 454
y 81 CysProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaVal 100
b 455 TGTCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCGCTGCGAGTG 514
y 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
b 515 GCAGTCCGCCTCTCCAAGGACCGATCCACTGCAGGTGCTGAGTCTGGCCACAGGGAAC 574
y 121 TrpPheSerAlaCysPheAspAsnPheThrGlu-AlaLeuAlaGluThrAlaCysArgG1 140
b 575 TGGTTCTCTGCCTGTTTCGACAACTTCAGAAAGCTCTCGTGAGACAGCCTGTAGGCA 634
y 140 nMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLe 160
b 635 GATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACAGGATCT 694
y 160 uAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPr 180
b 695 GGATGTCGTTGAATACAGAAACAGCCAGGAGCTTCGCATGCGGAACTCAAGTGGGCC 754
y 180 cCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSer 197

Db 755 -TGTCTCTCAGGCTC-CTGGTCTCCCTGAACTGTCTTGCCTGGGGAAGAGC 804

RESULT 11

BG386903

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 1007)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: NIH Intramural Sequencing Center

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM1306 row: c column: 02

High quality sequence stop: 763.

Location/Qualifiers

1. .1007

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4582969"

/tissue_type="adenocarcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_15"

/note="Organ: colon; Vector: pOTB7; Site 1: XhoI; Site 2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 240 a 291 c 293 g 183 t

ORIGIN

Alignment Scores:

Pred. No.: 4.53e-92 Length: 1007

Score: 949.50 Matches: 207

Percent Similarity: 84.17% Conservative: 11

Best Local Similarity: 79.92% Mismatches: 22

Query Match: 40.54% Indels: 19

DB: 10 Gaps: 2

US-09-607-745-2 (1-435) x BG386903 (1-1007)

QY 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20

Db 215 ATGGATCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAACCCCTGCGCAACCC 274

QY 21 ArgileProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSer 40

Db 275 CGTATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 334

QY 41 LeuAlaSerIleIleValValValIleLysValIleLeuAspLysTyrTyrPhe 60

Db 335 CTGGCGAGTATCATATTGTGTTGTCTCATCAAGGTGATCTGGATAAATACTACTTC 394

QY 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80


```
b 395 CTCTGGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGAC 454
y 81 CysProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaVal 100
b 455 TGTCCTCTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTGCAGTG 514
y 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
b 515 GCAGTCCGGCTCTCCAAGGACCGATCCACACTGCAGGTGGTGGACTCGGCCACAGGGAAC 574
y 121 TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
b 575 TGGTTCCTCGCTGTTTCGACAACCTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAG 634
y 141 MetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
b 635 ATGGGCTACAGCAGCAACCCACTTTCAGAGCTGTGGAGATTGGGCCAGAACAGGATCTG 694
y 161 AspValValGluIleThr-GluAsnSerGlnGluLeu-ArgMetArgAsnSerSer-Gly 179
b 695 GATGTCGTTGACATCACCAGAAACAGCCAGGAGCTTCCGAATGCGGAACCTCAAGTGGG 754
y 180 ProCysLeuSerGlySerLeuValSerLeuHisCysLeuAla-CysGlyLysSerLeuLy 199
b 755 CCTGCTCTCAGGCTCCCTGGTCTCCTTGACTGGTTTCGCCCTGTGGGAAGAGCTGAA 814
y 199 s-ThrProArgValValGlyGlyGluGluAlaSerVal---AspSerTrpProTrpGlnV 218
b 815 GAACCCCGTGTGTGCGCTTGTGGAAGAGGCCCCCAGCGAGAACATCAAGGCTTAGGGAAG 874
y 218 alSer-IleGlnTyrAspLysGlnHisValCys----- 228
b 875 AGACACAATCAGTACCAACAACACCCGAGTCTGTGAGAGAGAGACACTCGGGGACCCACG 934
y 229 -----GlyGlySerIleLeuAspProHisTrpValLeuThr 240
b 935 GGCACCTACAGGAAGAACCTCGTATCAGAGCACTACCGTGTTACA 979

BSULT 12
Q922656
OCUS
EFINITION
BQ922656
BQ922656
BQ922656.1 GI:22337687
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13891 row: f column: 06
High quality sequence stop: 599.
Location/Qualifiers
1. .920
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6395909"
```

```
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Co24"
/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 216 a 230 c 255 g 213 t 6 others
ORIGIN
Alignment Scores:
Pred. No.: 5.28e-86 Length: 920
Score: 893.00 Matches: 174
Percent Similarity: 75.95% Conservative: 25
Best Local Similarity: 66.41% Mismatches: 52
Query Match: 38.13% Indels: 12
DB: 13 Gaps: 2
US-09-607-745-2 (1-435) x BQ922656 (1-920)
QY 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
Db 9 TGCCTCTCAGGCTCCCTGGTTCCTTGGCGTGCCTTGACTGTGGAAAGAGCCTGAAGACT 68
QY 201 ProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
Db 69 CCTCGTGTGGTGGTGGGTGGAGGCCCTTGIGATTCTTGGCCGTGGCAGGTCAGCATC 128
QY 221 GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
Db 129 CAGTACAAACAGCAGCATGTCTGTGGTGGAGCATCCTGGATCCCCACTGGATCCTCACA 188
QY 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260
Db 189 GCAGCCCACTGCTTCAGGAAGTATCTTGATGTGTCAAGCTGGAAGGTGAGGCGAGGCTCA 248
QY 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsn 280
Db 249 AACATACTGGGTAACTCTCCATCCTTGCCTGTGGCAAGATCTTCATCGTGAACCAAT 308
QY 281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
Db 309 CCTCTGTACCCCAAGAGAGACATTGCCCTTGTAAAGCTGCAGATGCCACTCACATTC 368
QY 301 SerGlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThr 320
Db 369 TCAGGCTCAGTCAGGCCCATCTGCTGCCCTCTCTGTATGAGGTGCTTGTCCAGCCACA 428
QY 321 ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle 340
Db 429 CCAGTCTGGGTCACTTGGATGGGGCTTTACAGAAAGAAACCGAGGAAGATGTCTGACATG 488
QY 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyr 360
Db 489 CTACTGCANGCATCAGTCCAGGTTCATCGACAGCACACGGTGCATGCAGAGGATGCCTAC 548
QY 361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr 380
Db 549 GAAGGGGAAGTGACCGCTGAGATGCTGTGTGAGGTACCCACACAGGTGGCAGGACACC 608
QY 381 CysGln-----GlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrp 396
Db 609 TGCCAGGTGGGTCTCTCCGAAGACCCGAGGAGCTCTGTGAAGAGGGAATGCAAGTCCTAG 668
QY 397 HisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyVal 416
Db 669 AGAGATGTGAACCC-CTACAGGCTGGATGCCATGTGGAGGTTGTAGCGTCAAGTTGTG 727
QY 417 Tyr-----ThrLysValSerAlaTyrLeuAsnTrpIleTyrAsn 429
Db 728 ATAAAGGAAGGCATAATATGCTGTGAGAAAGTTAAANNAGGNATGGGTGGATAGGTGAT 787
QY 430 ValTrp 431
```

b 788 AGTTGG 793

RESULT 13
G966811
OCUS
EFINITION
CCESION
ERSION
EYWORDS
OURCE
ORGANISM

BG966811 689 bp mRNA linear EST 12-JUN-2001
602834306F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4989048 5',
mRNA sequence.
BG966811
BG966811.1 GI:14354448
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 689)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11002 row: k column: 01
High quality sequence stop: 680.
Location/Qualifiers
1. .689
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4989048"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Co24"
/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library."
ASE COUNT 160 a 189 c 188 g 152 t
RIGIN

Alignment Scores:
red. No.: 5.54e-86 Length: 689
core: 891.00 Matches: 165
ercent Similarity: 90.00% Conservative: 24
est Local Similarity: 78.57% Mismatches: 21
uery Match: 38.04% Indels: 2
B: 12 Gaps: 0

S-09-607-745-2 (1-435) x BG966811 (1-689)

Y 226 HisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAlaAlaHisCysPhe 245
|||||
b 2 CATGTCTGTGGGAGCATCCTGGATCCCACTGGATCCTCAGCAGCCCACTGCTTC 61
|||||

Y 246 ArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAspLysLeuGlySer 265
|||||
b 62 AGGAAGTATCTTGATGTGTCAAGCTGGAAGTCAAGGTCAGGCTCAACATCTGGGTAAC 121
|||||

Y 266 PheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnProMetTyrProLys 285
|||||
b 122 TCTCCATCCTTGCCTGTGGCCCAAGATCTTATCGTGAACCCATCTCTGTACCCCAA 181
|||||

Y 286 AspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArg 305
:::
b 182 GAGAAGGACATTGCCCTGTGTTAAGCTGCAGATGCCACTCACATTCTCAGGCTCAGTCAGG 241
|||||

Y 306 ProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrProLeuTrpIleIle 325
|||||

Db 242 CCCATCTGCCTGCCCTTCTCTGATGAGGTTGTTGCCAGCCACACCATCTGGTCATT 301

QY 326 GlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeuGlnAlaSer 345
|||||
Db 302 GGATGGGGCTTTACAGAAGAAACGGAGGAAGATGTCTGACATGCTACTGCAGGCATCA 361
|||||

QY 346 ValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGlnGlyGluValThr 365
|||||
Db 362 GTCCAGGTCAATCGACAGCACACGGTGTCAATGAGAGGATGCCTACGAAGGGAAGTGACC 421
|||||

QY 366 GluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCysGlnGlyAspSer 385
:::
Db 422 GCTGAGATGCTGTGTGCAGGTACCCACACAGGTGGCAAGACACCTGCCA-GGTGACAGT 480
|||||

QY 386 GlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValGlyIleValSerTrpGly 405
|||||
Db 481 GGTGGGCCCTTGTATGTACCATCTGTACAAAGTGGCAGGTAGTAGGCATCGTGAGCTGGGGC 540
|||||

QY 406 TyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsn 425
:::
Db 541 CATGGATGCGGGCGG-CCAAAGTACTCTCTGGAGTGTATATACCAAGGTCACTGCCTATCTCAAC 599
|||||

QY 426 TrpIleTyrAsnValTrpLysAlaGluLeu 435
|||||

Db 600 TGGATCTACAATGTTCCGGAAGTCTGAGATG 629
|||||

RESULT 14
BE616186 879 bp mRNA linear EST 24-AUG-2000
LOCUS 601278758F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3610695 5',
DEFINITION mRNA sequence.
ACCESSION BE616186
VERSION BE616186.1 GI:9897785
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 879)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM267 row: c column: 16
High quality sequence stop: 733.
Location/Qualifiers
1. .879
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3610695"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 39"
/note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies)."
BASE COUNT 182 a 248 c 274 g 175 t
ORIGIN

Alignment Scores:

```
red. No.:      9.18e-86      Length:      879
core:          890.50      Matches:      198
percent Similarity: 86.32%      Conservative: 4
est Local Similarity: 84.62%      Mismatches: 24
very Match:    38.02%      Indels:    11
B:            10          Gaps:        2
```

S-09-607-745-2 (1-435) x BB616186 (1-879)

```
Y      2 AspProAspSerAspGlnProLeuAsnSerLeuAspVallysProLeuArgLysProArg 21
b     198 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAACCCCTGCGAAACCCCGT 257

Y      22 IleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSerLeu 41
b     258 ATCCCCATGGAGACTTCAGAAAGGTGGGGATCCCCATCATCATAGCACTACTGAGCCCTG 317

Y      42 AlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
b     318 GCGAGTATCATCATTTGTTGCTCTCATCAAGGTGATTTCTGGATAAATACTACTTCTCTC 377

Y      62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
b     378 TCGGGGAGCCCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGGAGAGCTGGACTGT 437

Y      82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
b     438 CCCTTGGGGAGGAGGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTGCAGTGGCA 497

Y     102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
b     498 GTCCGCCTCTCCAGGACCGATCCACACTGCAGGTGTGGACTCGGC-ACAGGGAAGTGG 556

Y     122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
b     557 TTCTCTGCCTTTTCGACAACTTCACAGAGCTCTCGTGTGAGACAGC-TGTAGGCAGATG 615

Y     142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyPro-AspGlnAspLeuAs 161
b     616 GGCTACAGCAGCAAAACCACTTTACAGAGCTGTGGAGATTGGGCAGGACACAGGATCTGGA 675

Y     161 pValValGluIleThrGluAsnSerGlnGluLeu-ArgMetArgAsnSerSerGlyProC 181
b     676 TGTGTTGTAATCACAGAACACACAGCAGGAGCTTTTCGATCGGGAACCTCAAGTGGGCC-T 734

Y     181 YsLeuSerGlySerLeuVal-SerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
b     735 GTCTCTCAGGGTCCCTGGTCTTCCCTGCAGTGTGTGCTGTGGGAAAG--CTGAAACCC 791

Y     201 ProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
b     792 CCCCTGTTGGGGGGGGGGGGGGGGCC-----TGGGGGATATATATGGCGTGGGG 839

Y     221 GlnTyrAspLysGlnHisValCysGlyGlySerIle 232
b     840 GGACATTCCCGCCACAGCGCGCTGGGGGGGTCTGTG 875
```

RESULT 15

B142902

OCUS

EFINITION K-EST0196799 L11SNU354s1 Homo sapiens cDNA clone L11SNU354s1-20-D07

5', mRNA sequence.

CCESSION CB142902

ERSION CB142902.1 GI:28119586

EYWORDS EST.

OURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 515)

EFERENCE

AUTHORS

Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,

Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and

TITLE
JOURNAL
COMMENT

Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 20 row: D column: 07
High quality sequence stop: 515.

FEATURES

source

```
1..515
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L11SNU354s1-20-D07"
/sex="M"
/tissue_type="Liver"
/cell_type="Polygonal"
/cell_line="SNU-354"
/lab_host="Top10F/"
/clone_lib="L11SNU354s1"
/note="Organ: Liver; Vector: pCNS-D2; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including
EcoRI site by treatment of T4 RNA ligase and the first
strand cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 - 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and N(drf)14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNAs prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the subtracted cDNA
libraries were constructed by transformaion of the
remaining DNA into competent cells E. coli Top10F' with
electroporation method."
```

```
BASE COUNT 116 a 139 c 150 g 110 t
ORIGIN
```

Alignment Scores:

```
Pred. No.:      1.57e-85      Length:      515
Score:          885.00      Matches:      160
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    37.79%      Indels:    0
DB:            14          Gaps:        0
```

US-09-607-745-2 (1-435) x CB142902 (1-515)

QY

276 IleIleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGln 295

Db

2 ATCATGTAATTCAACCCCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAG 61

QY

296 PheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGlu 315

Db

62 TTCCCACTCACTTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTGATGAGGAG 121

```
Y 316 LeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGly 335
b 122 CTCACCTCCAGCCACCCCACTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAAATGGAGG 181

Y 336 LysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsn 355
b 182 AAGATGTCTGACATACTGCTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGCAAT 241

Y 356 AlaAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGlu 375
b 242 GCAGACGATGCGTACCAGGGGGGAAGTCACCGAGAAGATGATGTGTGAGGCATCCCGGAA 301

Y 376 GlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGln 395
b 302 GGGGGTGTGGACACCTGCCAGGGTGACAGTGTGGGCCCCCTGATGTACCAATCTGACCAG 361

Y 396 TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGly 415
b 362 TGGCATGTGTGGTGGGCATCGTTAGTTGGGGCTATGGCTGCGGGGGCCCGAGCACCCGAGGA 421

Y 416 ValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
b 422 GTATACACCAAGGTCTCAGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 481
```

earch completed: November 23, 2003, 12:05:54
ob time : 3113.27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

protein - protein search, using sw model
on: November 23, 2003, 07:38:25 ; Search time 42.9766 Seconds
(without alignments)
1078.451 Million cell updates/sec

itle: US-09-607-745-9
erfect score: 1600
equence: 1 MDSKGSSQKSRLLLLLVSN.....LNWYVWKAELSRHHHHH 292

oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

sarched: 1107863 seqs, 158726573 residues

tal number of hits satisfying chosen parameters: 1107863

inimum DB seq length: 0
aximum DB seq length: 2000000000
ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : A Geneseq 19Jun03: *
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT: *
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT: *
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT: *
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT: *
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT: *
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT: *
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT: *
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT: *
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT: *
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT: *
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT: *
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT: *
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT: *
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT: *
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT: *
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT: *
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT: *
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT: *
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT: *
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT: *
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT: *
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT: *
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT: *
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match %	Length	ID	Description
1	1600	100.0	292	23	Protease D-G catay
2	1266	79.1	317	22	Human protein sequ
3	1266	79.1	405	24	Human expressed pr
4	1266	79.1	405	24	Human expressed pr
5	1266	79.1	406	23	CJA8 sequence. Ho
6	1266	79.1	423	22	Human transmembran
7	1266	79.1	423	24	Human colorectal c
8	1266	79.1	423	24	Human expressed pr
9	1266	79.1	423	24	Human expressed pr

10	1266	79.1	423	24	ABU04923	Human expressed pr
11	1266	79.1	432	21	AAU99417	Human PRO1570 (UNQ
12	1266	79.1	432	22	AAU29188	Human PRO polypept
13	1266	79.1	432	22	AAB87581	Human PRO1570. Ho
14	1266	79.1	432	22	AAB66166	Protein of the inv
15	1266	79.1	432	23	ABG95906	Human secreted/tra
16	1266	79.1	432	23	AAU76535	Tumour-associated
17	1266	79.1	432	24	ABU71276	Human PRO1570 prot
18	1266	79.1	432	24	ABU71561	Human secreted pol
19	1266	79.1	432	24	ABU72007	Novel human secret
20	1266	79.1	432	24	ABU72164	Human PRO polypept
21	1266	79.1	432	24	ABU65733	Human secreted/tra
22	1266	79.1	432	24	ABU66066	Novel human secret
23	1266	79.1	432	24	ABU67570	Human secreted/tra
24	1266	79.1	432	24	ABU65428	Human PRO polypept
25	1266	79.1	432	24	ABU58564	Human PRO polypept
26	1266	79.1	432	24	ABU56100	Human secreted/tra
27	1266	79.1	432	24	ABU57095	Human PRO polypept
28	1266	79.1	432	24	ABU04921	Human expressed pr
29	1266	79.1	432	24	ABU04929	Human expressed pr
30	1266	79.1	432	24	ABU04934	Human expressed pr
31	1266	79.1	432	24	ABU10674	Human secreted/tra
32	1266	79.1	435	20	AAU06437	Human protease HUP
33	1266	79.1	435	22	AAU72558	Human seripancrin
34	1266	79.1	435	23	AAG78577	Human serine prote
35	1266	79.1	435	24	ABU04912	Human expressed pr
36	1266	79.1	435	24	ABU04930	Human expressed pr
37	1266	79.1	435	24	ABU04931	Human expressed pr
38	1266	79.1	437	23	AAG79359	CJA8 preferred seq
39	1266	79.1	437	24	ABU04915	Human expressed pr
40	1266	79.1	437	24	ABU04920	Human expressed pr
41	1266	79.1	437	24	ABU04935	Human expressed pr
42	1266	79.1	437	24	ABU04936	Human expressed pr
43	1264	79.0	233	18	AAW22986	Human serine prote
44	1253	78.3	492	22	AAU72559	Human seripancrin
45	1253	78.3	492	24	ABU04913	Human expressed pr

ALIGNMENTS

RESULT 1
AAG78578
ID AAG78578 standard; Protein; 292 AA.
XX
AC AAG78578;
XX
DT 07-MAY-2002 (first entry)
XX
DE Protease D-G cataytic domain fusion gene construct amino acid sequence.
XX
KW Serine protease; D-G; human; zymogen; enzyme; cytostatic;
KW antiinflammatory; dermatological; anticoagulation; cancer;
KW skin disorder; neuropathic pain; inflammatory disorder;
KW coagulation diathesis; thrombosis; laundry detergent; skin care;
KW gene therapy.
XX
OS Homo sapiens.
XX

Key	Location/Qualifiers
Peptide	1..27
Protein	/label= Prolactin_signal_sequence 28..292
Domain	/label= Mature_protease_D-G 60..284
Misc-difference	/note= "Protease D-G catalytic domain"
Region	/note= "Encoded by CTG"
WO200202011-A1.	/note= "6xHIS-TAG"

Y 51 KIVGGYALDVSWPQVSIQYDKQHVCGSILDPHWLTAAHCFRKHTDVFNWKVRAGSD 110
b 84 RVVGGEASVDSWPQVSIQYDKQHVCGSILDPHWLTAAHCFRKHTDVFNWKVRAGSD 143
Y 111 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPATP 170
b 144 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPATP 203
Y 171 LWIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAYQGEVTEKMMKAGIPEGGVDTTC 230
b 204 LWIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAYQGEVTEKMMKAGIPEGGVDTTC 263
Y 231 QGDSGGPLMYQSDQWHVVGVISWGYGCGGPSTPGVYTKVSAYLNWIYNVWKAE 284
b 264 QGDSGGPLMYQSDQWHVVGVISWGYGCGGPSTPGVYTKVSAYLNWIYNVWKAE 317

RESULT 3
BU04916
D ABU04916 standard; Protein; 405 AA.
X ABU04916;
C
X 29-JAN-2003 (first entry)
T Human expressed protein tag (EPT) #1582.
X
E Translational profiling; expressed protein tag; EPT; kinase;
W phosphatase; protease; protease inhibitor; transporter;
W cytoskeletal protein; receptor; transcription factor; cancer; MHC;
W major histocompatibility complex; myeloma; colon cancer;
W gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
W leukaemia.

X Homo sapiens.
S
X WO200278524-A2.
X 10-OCT-2002.
X 28-MAR-2002; 2002WO-US09671.
X 28-MAR-2001; 2001US-279495P.
R 21-MAY-2001; 2001US-292544P.
R 08-AUG-2001; 2001US-310801P.
R 01-OCT-2001; 2001US-326370P.
R 04-DEC-2001; 2001US-336780P.
R 20-FEB-2002; 2002US-358985P.
X (ZYCO-) ZYCOS INC.
A
X Chiciz RM, Tomlinson AJ, Urban RG;
I
X WPI; 2003-040607/03.
X New polypeptides (e.g. kinases, phosphatases, proteases, transporters, cytoskeletal proteins, receptors or transcription factors), useful for treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or leukemia -
S Example 2; SEQ ID No 1582; 134pp; English.
X
C The invention describes a purified polypeptide, which comprises a fragment of a kinase, phosphatase, protease, protease inhibitor, transporter, cytoskeletal protein, receptor or transcription factor.
C The polypeptide is useful as an immunogenic composition for eliciting in a mammal an immunogenic response directed against any of the purified polypeptide. The purified polypeptide, or the antibody that binds to this polypeptide, is useful for treating cancer. The polypeptide is also useful for identifying compounds that binds to a naturally processed class I or class II MHC-binding polypeptide. The polypeptides and polynucleotides are particularly useful for treating or preventing

CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma, lymphoma or leukaemia. These are also useful for screening agents for treating the above mentioned diseases. This sequence represents an expressed protein tag (EPT) isolated from human tissue for translational profiling.
CC Note: This sequence does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 405 AA;
Query Match 79.1%; Score 1266; DB 24; Length 405;
Best Local Similarity 97.4%; Pred. No. 6.6e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 51 KIVGGYALDVSWPQVSIQYDKQHVCGSILDPHWLTAAHCFRKHTDVFNWKVRAGSD 110
Db 172 RVVGGEASVDSWPQVSIQYDKQHVCGSILDPHWLTAAHCFRKHTDVFNWKVRAGSD 231
QY 111 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPATP 170
Db 232 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPATP 291
QY 171 LWIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAYQGEVTEKMMKAGIPEGGVDTTC 230
Db 292 LWIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAYQGEVTEKMMKAGIPEGGVDTTC 351
QY 231 QGDSGGPLMYQSDQWHVVGVISWGYGCGGPSTPGVYTKVSAYLNWIYNVWKAE 284
Db 352 QGDSGGPLMYQSDQWHVVGVISWGYGCGGPSTPGVYTKVSAYLNWIYNVWKAE 405

RESULT 4
ABU04937
ID ABU04937 standard; Protein; 405 AA.
XX
AC ABU04937;
DT 29-JAN-2003 (first entry)
XX Human expressed protein tag (EPT) #1603.
DE
XX Translational profiling; expressed protein tag; EPT; kinase; phosphatase; protease; protease inhibitor; transporter; cytoskeletal protein; receptor; transcription factor; cancer; MHC; major histocompatibility complex; myeloma; colon cancer; gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX Homo sapiens.
OS
XX WO200278524-A2.
PN
XX 10-OCT-2002.
XX 28-MAR-2002; 2002WO-US09671.
PF
XX 28-MAR-2001; 2001US-279495P.
PR 21-MAY-2001; 2001US-292544P.
PR 08-AUG-2001; 2001US-310801P.
PR 01-OCT-2001; 2001US-326370P.
PR 04-DEC-2001; 2001US-336780P.
PR 20-FEB-2002; 2002US-358985P.
XX (ZYCO-) ZYCOS INC.
PA Chiciz RM, Tomlinson AJ, Urban RG;
XX WPI; 2003-040607/03.
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters, cytoskeletal proteins, receptors or transcription factors), useful for treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma

T or leukemia -
X Example 2; SEQ ID No 1603; 134pp; English.
S
X
C The invention describes a purified polypeptide, which comprises a
C fragment of a kinase, phosphatase, protease, protease inhibitor,
C transporter, cytoskeletal protein, receptor or transcription factor.
C The polypeptide is useful as an immunogenic composition for eliciting
C in a mammal an immunogenic response directed against any of the purified
C polypeptide. The purified polypeptide, or the antibody that binds to
C this polypeptide, is useful for treating cancer. The polypeptide is
C also useful for identifying compounds that binds to a naturally
C processed class I or class II MHC-binding polypeptide. The polypeptides
C and polynucleotides are particularly useful for treating or preventing
C myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
C lymphoma or leukaemia. These are also useful for screening agents for
C treating the above mentioned diseases. This sequence represents an
C expressed protein tag (EPT) isolated from human tissue for translational
C profiling.
C Note: This sequence does not appear in the printed specification but was
C obtained in electronic format directly from WIPO at
C ftp.wipo.int/pub/published_pct_sequences.
X
Q Sequence 405 AA;

Query Match 79.1%; Score 1266; DB 24; Length 405;
Best Local Similarity 97.4%; Pred. No. 6.6e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Y 51 KIVGGYALDVDSWPQVSIQYDKQHVCGSILDPHWLTAHCFRKHTDVFNNKVRAGSD 110
b :|||||
172 RVVGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTAHCFRKHTDVFNNKVRAGSD 231

Y 111 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 170
b |||||||
232 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 291

Y 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCAGIPEGGVDT 230
b |||||||
292 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCAGIPEGGVDT 351

Y 231 QGDSGGLMYQSDQWHVVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAEL 284
b |||||||
352 QGDSGGLMYQSDQWHVVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAEL 405

ESULT 5
AG79358
D AAG79358 standard; Protein; 406 AA.
X
C AAG79358;
X
T 21-AUG-2002 (first entry)
X
E CJA8 sequence.
X
W Colorectal cancer; CGA7; CJA8; modulating protein; screening;
W drug candidate; vaccine.
X Homo sapiens.
X
H Key Location/Qualifiers
T Domain 5..32
T /note= "Transmembrane domain"
X
N US2002042067-A1.
X
D 11-APR-2002.
X
F 08-MAY-2001; 2001US-0851588.
X
R 17-AUG-2000; 2000US-0642252.
R 06-SEP-2000; 2000US-0656002.

XX (MACK/) MACK D.
PA (GISH/) GISH K C.
PA (WILS/) WILSON K E.
XX
PI Mack D, Gish KC, Wilson KE;
XX
DR WPI; 2002-453647/48.
DR N-PSDB; AAI72975.
XX
PT Screening drug candidates for treating colorectal cancer, comprises
PT determining the effect of the candidate on the expression profile gene
PT of CGA7 or CJA8 -
XX
PS Disclosure; Fig 6; 40pp; English.
XX
CC The sequences given in AAG79356-59 show the colorectal cancer
CC proteins, CGA7 and CJA8. The CGA7 and CJA8 proteins are colorectal
CC cancer modulating proteins and have been mapped to chromosomes 2 (CGA7)
CC and 11 (CJA8). These sequences may be used in the method of the
CC invention for screening drug candidates. The method comprises adding
CC a drug candidate to a cell that expresses an expression profile gene
CC encoding CGA7, CJA8 or fragments and determining the effect of the
CC drug candidate on the expression of the expression profile gene. The new
CC methods are used to screen bioactive agents for the ability to bind to
CC or modulate the activity of CGA7 or CJA8 and evaluate the effect of a
CC candidate colorectal cancer drug. An antibody to CGA7 or CJA8 can inhibit
CC the activity of CGA7 or CJA8, respectively, and is used to screen for
CC an agent that can interfere with the binding of CGA7 or CJA8 to the
CC antibody. The antibody can be used to treat colorectal cancer. The
CC antibody or a fragment of it is used to localize a therapeutic group to
CC a colorectal cancer tissue, where the therapeutic group is a cytotoxic
CC agent or a radioisotope. Antisense molecules are used to inhibit
CC colorectal cancer in a cell. Nucleic acid segments encoding CGA7 or
CC CJA8 are used in a biochip. CGA7, CJA8 or a nucleic acid encoding it
CC are used to elicit an immune response. CGA7 or CJA8 is used to
CC determine the prognosis of an individual with colorectal cancer.
CC Nucleic acid encoding CGA7 or CJA8 can be used in vaccines.
XX
SQ Sequence 406 AA;

Query Match 79.1%; Score 1266; DB 23; Length 406;
Best Local Similarity 97.4%; Pred. No. 6.6e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 51 KIVGGYALDVDSWPQVSIQYDKQHVCGSILDPHWLTAHCFRKHTDVFNNKVRAGSD 110
Db :|||||
173 RVVGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTAHCFRKHTDVFNNKVRAGSD 232

QY 111 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 170
Db |||||||
233 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 292

QY 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCAGIPEGGVDT 230
Db |||||||
293 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCAGIPEGGVDT 352

QY 231 QGDSGGLMYQSDQWHVVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAEL 284
Db |||||||
353 QGDSGGLMYQSDQWHVVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAEL 406

RESULT 6
AAE06944
ID AAE06944 standard; Protein; 423 AA.
XX
AC AAE06944;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human transmembrane protease serine 4 (TMPRSS4) protein.
XX
KW Human; transmembrane serine protease; membrane-type serine protease;

Y 231 QGDSGGPLMYQSDQWHVVGIVSWGCGGPSTPGVYTKVSAYLNWIYNVWKAE 284
b 370 QGDSGGPLMYQSDQWHVVGIVSWGCGGPSTPGVYTKVSAYLNWIYNVWKAE 423
RESULT 8
BU04918
D ABU04918 standard; Protein; 423 AA.
X
C ABU04918;
X
T 29-JAN-2003 (first entry)
X Human expressed protein tag (EPT) #1584.
E
X Translational profiling; expressed protein tag; EPT; kinase;
W phosphatase; protease; protease inhibitor; transporter;
W cytoskeletal protein; receptor; transcription factor; cancer; MHC;
W major histocompatibility complex; myeloma; colon cancer;
W gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
W leukaemia.
X Homo sapiens.
S
X
X WO200278524-A2.
N
X
D 10-OCT-2002.
X
X 28-MAR-2002; 2002WO-US09671.
F
X 28-MAR-2001; 2001US-279495P.
R 21-MAY-2001; 2001US-292544P.
R 08-AUG-2001; 2001US-310801P.
R 01-OCT-2001; 2001US-326370P.
R 04-DEC-2001; 2001US-336780P.
R 20-FEB-2002; 2002US-358985P.
X
X (ZYCO-) ZYCOS INC.
A
X Chiciz RM, Tomlinson AJ, Urban RG;
X WPI; 2003-040607/03.
R
X New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
T cytoskeletal proteins, receptors or transcription factors), useful for
T treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
T or leukemia
X
S Example 2; SEQ ID No 1584; 134pp; English.
X
C The invention describes a purified polypeptide, which comprises a
C fragment of a kinase, phosphatase, protease, protease inhibitor,
C transporter, cytoskeletal protein, receptor or transcription factor.
C The polypeptide is useful as an immunogenic composition for eliciting
C in a mammal an immunogenic response directed against any of the purified
C polypeptide. The purified polypeptide, or the antibody that binds to
C this polypeptide, is useful for treating cancer. The polypeptide is
C also useful for identifying compounds that binds to a naturally
C processed class I or class II MHC-binding polypeptide. The polypeptides
C and polynucleotides are particularly useful for treating or preventing
C myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
C lymphoma or leukaemia. These are also useful for screening agents for
C treating the above mentioned diseases. This sequence represents an
C expressed protein tag (EPT) isolated from human tissue for translational
C profiling.
C Note: This sequence does not appear in the printed specification but was
C obtained in electronic format directly from WIPO at
C ftp.wipo.int/pub/published_pct_sequences.
X
Q Sequence 423 AA;

Query Match 79.1%; Score 1266; DB 24; Length 423;
Best Local Similarity 97.4%; Pred. No. 7e-112;

Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 51 KIVGGYALDVDSWPQVSIQYDKQHVCGSILDPHWLTAACFRKHTDVFNWKVRAGSD 110
Db 190 RVVGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTAACFRKHTDVFNWKVRAGSD 249
QY 111 KLGSPFLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPATP 170
Db 250 KLGSPFLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPATP 309
QY 171 LMIIGWFTKQNGKMSDILLOASVQVIDSTECNADDAAYQGEVTERKMCAGIPEGGVDT 230
Db 310 LMIIGWFTKQNGKMSDILLOASVQVIDSTECNADDAAYQGEVTERKMCAGIPEGGVDT 369
QY 231 QGDSGGPLMYQSDQWHVVGIVSWGCGGPSTPGVYTKVSAYLNWIYNVWKAE 284
Db 370 QGDSGGPLMYQSDQWHVVGIVSWGCGGPSTPGVYTKVSAYLNWIYNVWKAE 423
RESULT 9
ABU04919
ID ABU04919 standard; Protein; 423 AA.
XX
AC ABU04919;
XX
DT 29-JAN-2003 (first entry)
XX
DE Human expressed protein tag (EPT) #1585.
XX
KW Translational profiling; expressed protein tag; EPT; kinase;
KW phosphatase; protease; protease inhibitor; transporter;
KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer;
KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
KW leukaemia.
XX
OS Homo sapiens.
XX WO200278524-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WO-US09671.
XX
PR 28-MAR-2001; 2001US-279495P.
PR 21-MAY-2001; 2001US-292544P.
PR 08-AUG-2001; 2001US-310801P.
PR 01-OCT-2001; 2001US-326370P.
PR 04-DEC-2001; 2001US-336780P.
PR 20-FEB-2002; 2002US-358985P.
XX
PA (ZYCO-) ZYCOS INC.
XX
PI Chiciz RM, Tomlinson AJ, Urban RG;
XX
DR WPI; 2003-040607/03.
XX
PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
PT or leukemia
XX
PS Example 2; SEQ ID No 1585; 134pp; English.
XX
CC The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor.
CC The polypeptide is useful as an immunogenic composition for eliciting
CC in a mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to
CC this polypeptide, is useful for treating cancer. The polypeptide is
CC also useful for identifying compounds that binds to a naturally
CC processed class I or class II MHC-binding polypeptide. The polypeptides

and polynucleotides are particularly useful for treating or preventing myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma, lymphoma or leukaemia. These are also useful for screening agents for treating the above mentioned diseases. This sequence represents an expressed protein tag (EPT) isolated from human tissue for translational profiling.
Note: This sequence does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 423 AA;
Query Match 79.1%; Score 1266; DB 24; Length 423;
Best Local Similarity 97.4%; Pred. No. 7e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Y 51 KIVGGYALDVDSWPWQVSIQYDKQHVCGGSILDPHWLTAACHFRKHTDVFNNKVRAGSD 110
b 190 RVVGGEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTAACHFRKHTDVFNNKVRAGSD 249
Y 111 KLGSPSLAVAKIIIEFNPMYKNDIALMKLQFPLTSGTVRPICLPFFDELTPTATP 170
b 250 KLGSPSLAVAKIIIEFNPMYKNDIALMKLQFPLTSGTVRPICLPFFDELTPTATP 309
Y 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCAGIPEGVDTC 230
b 310 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCAGIPEGVDTC 369
Y 231 QGDSGGPLMYQSDQWHVVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 284
b 370 QGDSGGPLMYQSDQWHVVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 423

RESULT 10
BU04923
D ABU04923 standard; Protein; 423 AA.
X ABU04923;
C 29-JAN-2003 (first entry)
T Human expressed protein tag (EPT) #1589.
X Translational profiling; expressed protein tag; EPT; kinase;
E phosphatase; protease; protease inhibitor; transporter;
X cytoskeletal protein; receptor; transcription factor; cancer; MHC;
W major histocompatibility complex; myeloma; colon cancer;
W gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
W leukaemia.
X Homo sapiens.
S WO200278524-A2.
X 10-OCT-2002.
D 28-MAR-2002; 2002WO-US09671.
F 28-MAR-2001; 2001US-279495P.
X 21-MAY-2001; 2001US-292544P.
R 08-AUG-2001; 2001US-310801P.
R 01-OCT-2001; 2001US-326370P.
R 04-DEC-2001; 2001US-336780P.
R 20-FEB-2002; 2002US-358985P.
X (ZYCO-) ZYCOS INC.
A Chiciz RM, Tomlinson AJ, Urban RG;
X WPI; 2003-040607/03.
R New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
T cytoskeletal proteins, receptors or transcription factors), useful for

treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or leukemia -
XX Example 2; SEQ ID No 1589; 134pp; English.
PS The invention describes a purified polypeptide, which comprises a fragment of a kinase, phosphatase, protease, protease inhibitor, transporter, cytoskeletal protein, receptor or transcription factor.
CC The polypeptide is useful as an immunogenic composition for eliciting in a mammal an immunogenic response directed against any of the purified polypeptide. The purified polypeptide, or the antibody that binds to this polypeptide, is useful for treating cancer. The polypeptide is also useful for identifying compounds that binds to a naturally processed class I or class II MHC-binding polypeptide. The polypeptides and polynucleotides are particularly useful for treating or preventing myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma, lymphoma or leukaemia. These are also useful for screening agents for treating the above mentioned diseases. This sequence represents an expressed protein tag (EPT) isolated from human tissue for translational profiling.
CC Note: This sequence does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 423 AA;
Query Match 79.1%; Score 1266; DB 24; Length 423;
Best Local Similarity 97.4%; Pred. No. 7e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 51 KIVGGYALDVDSWPWQVSIQYDKQHVCGGSILDPHWLTAACHFRKHTDVFNNKVRAGSD 110
b 190 RVVGGEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTAACHFRKHTDVFNNKVRAGSD 249
QY 111 KLGSPSLAVAKIIIEFNPMYKNDIALMKLQFPLTSGTVRPICLPFFDELTPTATP 170
b 250 KLGSPSLAVAKIIIEFNPMYKNDIALMKLQFPLTSGTVRPICLPFFDELTPTATP 309
QY 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCAGIPEGVDTC 230
b 310 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCAGIPEGVDTC 369
QY 231 QGDSGGPLMYQSDQWHVVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 284
b 370 QGDSGGPLMYQSDQWHVVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 423

RESULT 11
AA999417
ID AA999417 standard; Protein; 432 AA.
XX AA999417;
AC AA999417;
XX 08-AUG-2000 (first entry)
DT Human PRO1570 (UNQ776) amino acid sequence SEQ ID NO:275.
DE Human; PRO polypeptide; membrane bound protein; receptor; diagnosis; transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
KW Homo sapiens.
XX WO200012708-A2.
PN 09-MAR-2000.
PD 01-SEP-1999; 99WO-US20111.
XX 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.

R 02-SEP-1998; 98US-0098843.
R 09-SEP-1998; 98US-0099536.
R 09-SEP-1998; 98US-0099596.
R 09-SEP-1998; 98US-0099598.
R 09-SEP-1998; 98US-0099602.
R 09-SEP-1998; 98US-0099642.
R 10-SEP-1998; 98US-0099741.
R 10-SEP-1998; 98US-0099754.
R 10-SEP-1998; 98US-0099763.
R 10-SEP-1998; 98US-0099792.
R 10-SEP-1998; 98US-0099808.
R 10-SEP-1998; 98US-0099812.
R 10-SEP-1998; 98US-0099815.
R 10-SEP-1998; 98US-0099816.
R 15-SEP-1998; 98US-0100385.
R 15-SEP-1998; 98US-0100388.
R 15-SEP-1998; 98US-0100390.
R 16-SEP-1998; 98US-0100584.
R 16-SEP-1998; 98US-0100627.
R 16-SEP-1998; 98US-0100661.
R 16-SEP-1998; 98US-0100662.
R 16-SEP-1998; 98US-0100664.
R 17-SEP-1998; 98US-0100683.
R 17-SEP-1998; 98US-0100684.
R 17-SEP-1998; 98US-0100710.
R 17-SEP-1998; 98US-0100711.
R 17-SEP-1998; 98US-0100919.
R 17-SEP-1998; 98US-0100930.
R 18-SEP-1998; 98US-0100848.
R 18-SEP-1998; 98US-0100849.
R 18-SEP-1998; 98US-0101014.
R 18-SEP-1998; 98US-0101068.
R 18-SEP-1998; 98US-0101071.
R 22-SEP-1998; 98US-0101279.
R 23-SEP-1998; 98US-0101471.
R 23-SEP-1998; 98US-0101472.
R 23-SEP-1998; 98US-0101474.
R 23-SEP-1998; 98US-0101475.
R 23-SEP-1998; 98US-0101476.
R 23-SEP-1998; 98US-0101477.
R 23-SEP-1998; 98US-0101479.
R 24-SEP-1998; 98US-0101738.
R 24-SEP-1998; 98US-0101741.
R 24-SEP-1998; 98US-0101743.
R 24-SEP-1998; 98US-0101915.
R 24-SEP-1998; 98US-0101916.
R 29-SEP-1998; 98US-0102207.
R 29-SEP-1998; 98US-0102240.
R 29-SEP-1998; 98US-0102307.
R 29-SEP-1998; 98US-0102330.
R 29-SEP-1998; 98US-0102331.
R 30-SEP-1998; 98US-0102484.
R 30-SEP-1998; 98US-0102487.
R 30-SEP-1998; 98US-0102487.
R 30-SEP-1998; 98US-0102570.
R 30-SEP-1998; 98US-0102571.
R 01-OCT-1998; 98US-0102684.
R 01-OCT-1998; 98US-0102687.
R 02-OCT-1998; 98US-0102965.
R 06-OCT-1998; 98US-0103258.
R 06-OCT-1998; 98US-0103449.
R 07-OCT-1998; 98US-0103314.
R 07-OCT-1998; 98US-0103315.
R 07-OCT-1998; 98US-0103328.
R 07-OCT-1998; 98US-0103395.
R 07-OCT-1998; 98US-0103396.
R 07-OCT-1998; 98US-0103401.
R 08-OCT-1998; 98US-0103633.
R 08-OCT-1998; 98US-0103678.
R 08-OCT-1998; 98US-0103679.
R 08-OCT-1998; 98US-0103711.
R 14-OCT-1998; 98US-0104257.
R 20-OCT-1998; 98US-0104987.
R 20-OCT-1998; 98US-0105000.

PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105593.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0108500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108948.
PR 18-NOV-1998; 98US-0108949.
PR 18-NOV-1998; 98US-0108950.
PR 18-NOV-1998; 98US-0108951.
PR 18-NOV-1998; 98US-0108952.
PR 18-NOV-1998; 98US-0108958.
PR 18-NOV-1998; 98US-0108904.
XX
PA
XX
XX
PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX
DR WPI; 2000-237871/20.
DR N-PSDB; AAA37099.
XX
XX
PT New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
XX small molecule inhibitors of the relevant receptor/ligand interactions
PS Claim 12; Fig 156; 773pp; English.
XX
CC AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC transmembrane and receptor PRO proteins can be used for screening of
CC potential peptide or small molecule inhibitors of the relevant
CC receptor/ligand interactions. The polypeptides and nucleotide sequences
CC encoding then have various industrial applications, including uses as
CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC PCR primers and hybridisation probes used in the isolation of the PRO
CC polypeptides from the present invention.
XX
SQ Sequence 432 AA;

Query Match 79.1%; Score 1266; DB 21; Length 432;
Best Local Similarity 97.4%; Pred. No. 7.le-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Y 51 KIVGYALDVSWPWQVSIQYDKQHVCVCGSILDPHWLTAAHCFRKHTDVFNWKVRAGSD 110
b ::::|||||
199 RVVGGEASVDSWPWQVSIQYDKQHVCVCGSILDPHWLTAAHCFRKHTDVFNWKVRAGSD 258
Y 111 KLGSPFLAVAKIIIEFNPMYPKNDIALMKLOFPLTFSGTVRPICLPFFDEELTPATP 170
b :|||
259 KLGSPFLAVAKIIIEFNPMYPKNDIALMKLOFPLTFSGTVRPICLPFFDEELTPATP 318
Y 171 LWIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCAGIPEGGVDTC 230
b :|||
319 LWIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCAGIPEGGVDTC 378
Y 231 QGDSGGPLMYQSDQHVHVGVISWGYGCGPSTPGVYTKVSAYLNWIYNVWKAEL 284
b :|||
379 QGDSGGPLMYQSDQHVHVGVISWGYGCGPSTPGVYTKVSAYLNWIYNVWKAEL 432

RESULT 12
AAU29188 standard; Protein; 432 AA.
AAU29188;
18-DEC-2001 (first entry)
Human PRO polypeptide sequence #165.

PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
Homo sapiens.
WO200168848-A2.
20-SEP-2001.
28-FEB-2001; 2001WO-US06520.

01-MAR-2000; 2000WO-US05601.
02-MAR-2000; 2000WO-US05841.
03-MAR-2000; 2000US-187202P.
06-MAR-2000; 2000US-186968P.
14-MAR-2000; 2000US-189320P.
14-MAR-2000; 2000US-189328P.
15-MAR-2000; 2000WO-US06884.
21-MAR-2000; 2000US-190828P.
21-MAR-2000; 2000US-191007P.
21-MAR-2000; 2000US-191048P.
21-MAR-2000; 2000US-191314P.
28-MAR-2000; 2000US-192655P.
29-MAR-2000; 2000US-193032P.
29-MAR-2000; 2000US-193053P.
30-MAR-2000; 2000WO-US08439.
04-APR-2000; 2000US-19449P.
04-APR-2000; 2000US-194647P.
11-APR-2000; 2000US-195975P.
11-APR-2000; 2000US-196000P.
11-APR-2000; 2000US-196187P.
11-APR-2000; 2000US-196690P.
11-APR-2000; 2000US-196820P.
18-APR-2000; 2000US-198121P.
18-APR-2000; 2000US-198585P.
25-APR-2000; 2000US-199397P.
25-APR-2000; 2000US-199550P.
25-APR-2000; 2000US-199654P.
03-MAY-2000; 2000US-201516P.
17-MAY-2000; 2000WO-US13705.
22-MAY-2000; 2000WO-US14042.
30-MAY-2000; 2000WO-US14941.
02-JUN-2000; 2000WO-US15264.

PR 05-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-0644848.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX (GETH) GENENTECH INC.
PA Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
XX Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
PI WPI; 2001-602746/68.
XX N-PSDB; AAS46089.
DR Novel nucleic acids encoding PRO polypeptides, used to diagnose the
XX presence of tumours, such as prostate and breast tumours, in mammals and
PT to screen for modulators of the compounds -
PT Claim 11; Fig 330; 774pp; English.
XX Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
CC The PRO polypeptides and their associated nucleic acids can be used to
CC detect the presence of a tumour in a mammal by comparing the level of
CC expression of a PRO polypeptide in a test sample of cells from the animal
CC and a control sample of normal cells, whereby a higher level of
CC expression in the test sample indicates the presence of a tumour in the
CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
CC and rabbits but are preferably human. The polypeptides can be used to
CC stimulate tumour necrosis factor (TNF) alpha release from human blood,
CC when contacted with it. A specific polypeptide can be used to stimulate
CC the proliferation or differentiation of chondrocyte cells. The PRO
CC proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX Sequence 432 AA;
SQ

Query Match 79.1%; Score 1266; DB 22; Length 432;
Best Local Similarity 97.4%; Pred. No. 7.1e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 51 KIVGYALDVSWPWQVSIQYDKQHVCVCGSILDPHWLTAAHCFRKHTDVFNWKVRAGSD 110
Db ::::|||||
199 RVVGGEASVDSWPWQVSIQYDKQHVCVCGSILDPHWLTAAHCFRKHTDVFNWKVRAGSD 258
QY 111 KLGSPFLAVAKIIIEFNPMYPKNDIALMKLOFPLTFSGTVRPICLPFFDEELTPATP 170
Db :|||
259 KLGSPFLAVAKIIIEFNPMYPKNDIALMKLOFPLTFSGTVRPICLPFFDEELTPATP 318
QY 171 LWIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCAGIPEGGVDTC 230
Db :|||
319 LWIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCAGIPEGGVDTC 378
QY 231 QGDSGGPLMYQSDQHVHVGVISWGYGCGPSTPGVYTKVSAYLNWIYNVWKAEL 284
Db :|||
379 QGDSGGPLMYQSDQHVHVGVISWGYGCGPSTPGVYTKVSAYLNWIYNVWKAEL 432

RESULT 13
AAB87581
ID AAB87581 standard; Protein; 432 AA.
XX
AC AAB87581;
XX
DT 15-MAY-2001 (first entry)
XX
DE Human PRO1570.
XX
KW Human; PRO protein; mapping.

X S Homo sapiens.
X X WO200116318-A2.
X N 08-MAR-2001.
X X 24-AUG-2000; 2000WO-US23328.
X R 01-SEP-1999; 99WO-US20111.
X R 15-SEP-1999; 99WO-US21090.
X R 07-DEC-1999; 99US-0169495.
X R 09-DEC-1999; 99US-0170262.
X R 11-JAN-2000; 2000US-0175481.
X R 18-FEB-2000; 2000WO-US04341.
X R 18-FEB-2000; 2000WO-US04342.
X R 22-FEB-2000; 2000WO-US04414.
X R 01-MAR-2000; 2000WO-US05601.
X R 03-MAR-2000; 2000US-0187202.
X R 25-APR-2000; 2000US-0199397.
X R 22-MAY-2000; 2000WO-US14042.
X R 05-JUN-2000; 2000US-0209832.
X X (GETH) GENENTECH INC.

X X Baton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
X I Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
X R WPI; 2001-183260/18.
X R N-PSDB; AAF92113.

X T Eighty four nucleic acids encoding PRO polypeptides, useful in
X T molecular biology, including use as hybridization probes, and in
X T chromosome and gene mapping.

X S Claim 12; Fig 112; 278pp; English.

X C The present sequence is a human PRO polypeptide (secreted and
X C transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
X C anti-PRO antibodies are useful for preparation of a medicament useful in
X C the treatment of a condition which is responsive to the PRO protein,
X C agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
X C employed as molecular weight markers for protein electrophoresis. The PRO
X C coding sequence has applications in molecular biology, including use as
X C hybridisation probes, and in chromosome and gene mapping.

X Q Sequence 432 AA;

Query Match 79.1%; Score 1266; DB 22; Length 432;
Best Local Similarity 97.4%; Pred. No. 7.1e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Y 51 KIVGGYALDVDSWPWQVSIQYDKQHVCGGSILDPHWLTAACHFRKHTDVFNNKVRAGSD 110
b ::::|||||
199 RVVGGEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTAACHFRKHTDVFNNKVRAGSD 258
Y 111 KLGSPSLAVAKIIIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPATP 170
b :|||||
259 KLGSPSLAVAKIIIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPATP 318
Y 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMCAGIPEGGVDT 230
b :|||||
319 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMCAGIPEGGVDT 378
Y 231 QGDSGGPLMYQSDQWHVVGVIVSWGVCGBPSTPGVYTKVSAYLNWIYNWKAE 284
b :|||||
379 QGDSGGPLMYQSDQWHVVGVIVSWGVCGBPSTPGVYTKVSAYLNWIYNWKAE 432

RESULT 14
AB66166
D AAB66166 standard; protein; 432 AA.

AC AAB66166;
XX 02-APR-2001 (first entry)
XX Protein of the invention #78.
DE Secreted; transmembrane; gene therapy.
XX Unidentified.
XX WO200078961-A1.
PD 28-DEC-2000.
XX 18-FEB-2000; 2000WO-US04342.
XX 23-JUN-1999; 99US-0141037.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 16-DEC-1999; 99WO-US30095.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
XX (GETH) GENENTECH INC.

XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
XX Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
XX Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
XX Watanabe CK, Williams PM, Wood WI;
XX WPI; 2001-071395/08.

XX Secreted and transmembrane proteins and nucleic acids designated PRO,
XX useful as hybridization probes, in chromosome and gene mapping and gene
XX therapy -

XX Claim 1; Fig 156; 787pp; English.

XX The present invention relates to secreted and transmembrane proteins.
XX These proteins and the DNA encoding them may be used as hybridization
XX probes, in chromosome and gene mapping and in the generation of
XX anti-sense RNA and DNA. They may also be used to generate either
XX transgenic animals or knockout animals which are in turn useful for
XX development and screening of therapeutically useful reagents.
XX The nucleic acids may also be used in gene therapy.

SQ Sequence 432 AA;

Query Match 79.1%; Score 1266; DB 22; Length 432;
Best Local Similarity 97.4%; Pred. No. 7.1e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 51 KIVGGYALDVDSWPWQVSIQYDKQHVCGGSILDPHWLTAACHFRKHTDVFNNKVRAGSD 110
Db ::::|||||
199 RVVGGEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTAACHFRKHTDVFNNKVRAGSD 258
QY 111 KLGSPSLAVAKIIIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPATP 170
Db :|||||
259 KLGSPSLAVAKIIIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPATP 318
QY 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMCAGIPEGGVDT 230
Db :|||||
319 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMCAGIPEGGVDT 378
QY 231 QGDSGGPLMYQSDQWHVVGVIVSWGVCGBPSTPGVYTKVSAYLNWIYNWKAE 284
Db :|||||
379 QGDSGGPLMYQSDQWHVVGVIVSWGVCGBPSTPGVYTKVSAYLNWIYNWKAE 432

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - protein search, using sw model

un on: November 23, 2003, 07:47:10 ; Search time 14.8611 Seconds
(without alignments)
831.351 Million cell updates/sec

itle: US-09-607-745-9
erfect score: 1600
equence: 1 MDSKGSSQKSRLLLLLVSN.....LNWIYNVWKAEISRHHHHH 292

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 328717 seqs, 42310858 residues

otal number of hits satisfying chosen parameters: 328717

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match %	Length	ID	Description
1	1266	79.1	423	4	US-09-656-002-2
2	1266	79.1	435	3	US-09-008-271A-6
3	765	47.8	316	4	US-09-387-375-9
4	754.5	47.2	315	4	US-09-386-653A-9
5	746	46.6	328	4	US-09-386-642-11
6	740.5	46.3	327	4	US-09-386-629-8
7	672.5	42.0	289	4	US-09-386-642-14
8	645	40.3	288	4	US-09-386-642-13
9	601.5	37.6	306	4	US-09-386-642-12
10	593	37.1	319	4	US-09-386-642-12
11	574	35.9	284	4	US-09-386-642-54
12	526.5	32.9	418	1	US-08-508-448C-25
13	526.5	32.9	418	4	US-09-370-838-82
14	526.5	32.9	418	4	US-09-370-838-83
15	524.5	32.8	418	4	US-09-370-838-62
16	522.5	32.7	232	1	US-08-508-448C-19
17	507	31.7	454	3	US-09-518-046-2
18	506	31.6	492	3	US-09-342-749-2
19	506	31.6	492	4	US-09-691-840-2
20	505.5	31.6	376	4	US-09-820-002-2
21	505.5	31.6	417	4	US-09-820-002-4
22	499.5	31.2	256	2	US-09-027-337-3
23	499.5	31.2	256	4	US-09-644-600-3
24	497.5	31.1	255	3	US-08-944-483-67
25	497	31.1	283	3	US-08-807-151-1
26	497	31.1	283	4	US-09-478-957-1
27	491.5	30.7	416	2	US-09-000-846-2

28	489	30.6	638	2	US-08-681-151-3	Sequence 3, Appli
29	482	30.1	285	4	US-09-023-942A-26	Sequence 26, Appl
30	477	29.8	248	3	US-08-944-483-63	Sequence 63, Appl
31	463.5	29.0	246	2	US-08-978-404B-44	Sequence 44, Appl
32	463.5	29.0	250	3	US-08-944-483-68	Sequence 68, Appl
33	461.5	28.8	407	4	US-09-734-675-4	Sequence 4, Appli
34	456.5	28.5	314	4	US-09-023-942A-6	Sequence 6, Appli
35	455	28.4	284	4	US-09-387-375-7	Sequence 7, Appli
36	453.5	28.3	247	2	US-08-956-267A-2	Sequence 2, Appli
37	451.5	28.2	314	3	US-09-008-271A-3	Sequence 3, Appli
38	450.5	28.2	312	4	US-09-023-942A-4	Sequence 4, Appli
39	444	27.8	855	2	US-09-027-337-2	Sequence 2, Appli
40	444	27.8	855	4	US-09-644-600-2	Sequence 2, Appli
41	443	27.7	790	4	US-08-991-761A-13	Sequence 13, Appl
42	442.5	27.7	290	4	US-09-386-653A-7	Sequence 7, Appli
43	441.5	27.6	812	1	US-08-248-629A-1	Sequence 1, Appli
44	441.5	27.6	812	1	US-08-451-932-1	Sequence 1, Appli
45	441.5	27.6	812	1	US-08-452-260-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-656-002-2
; Sequence 2, Application US/09656002
; Patent No. 6455668
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND A
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-69108/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/656,002
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-656-002-2

Query Match		79.1%;	Score 1266;	DB 4;	Length 423;
Best Local Similarity		97.4%;	Pred. No. 5.5e-128;		
Matches 228;		Conservative	2;	Mismatches	4;
				Indels	0;
				Gaps	0;
QY	51	KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHDTVENWKRAGSD	110		
		: : :			
Db	190	RVVGGEASVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHDTVENWKRAGSD	249		
QY	111	KLGSFSLAVAKIIIEFNMPYKNDIALMKLQFLTFSGTVRPICLPFFDELTPTATP	170		
		: : :			
Db	250	KLGSFSLAVAKIIIEFNMPYKNDIALMKLQFLTFSGTVRPICLPFFDELTPTATP	309		
QY	171	LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMCAGIPEGGVDTG	230		
		: : :			
Db	310	LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMCAGIPEGGVDTG	369		
QY	231	QGDSGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAEI	284		
		: : :			
Db	370	QGDSGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAEI	423		

RESULT 2
US-09-008-271A-6

sequence 9, Application US/09387375

```

Query Match      47.2%; Score 754.5; DB 4; Length 315;
Best Local Similarity 48.4%; Pred. No. 4.6e-73;
Matches 156; Conservative 46; Mismatches 83; Indels 37; Gaps 10;

Y 1 MDSKGSQKSRLLLLLVSNLLLCQGVVSDYKDDDDVDAAALAAPFDDDDKIIVGGYALDV 60
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
b 1 MDSKGSQKSRLLLLLVSNLLLCQGVVSDYKDDDDVDAAALAAPFDDDDKIIVGGYALEE 60
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Y 61 DSWPQVSIQYDKQHVCVGGSIIDPHWVLTAAHCFRKHTDVFNMWVRAGSDKL-GSFPSLA 119
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
b 61 GEWPQVSIQRNGSHFCGGSLIAEQWVLTAAHCFRNTSETSLYQVLLGARQLVQPGPHAM 120
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Y 120 VAKIIIEEFNMY--PKNDIALMKLOFPLTFSGTVRPICLPFFDEELTPATPL--WII 174
  ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
b 121 YARVRQVESNPYQGTASSADVALVELEAPVPFTNYILPVCLP--DPSVIFETGNCWVVT 178
  ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Y 175 GWGFTKQNGGKMSD-----ILLQASVQVIDSTRCN-----ADDAYQGE-VTEKMMCAG 221
  ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
b 179 GWGSPSE-----EDLLPEPRILQKLAVPIIDTPKCNLLYSKDTFGYQPKTIKNDMLCAG 233
  ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Y 222 IPEGGVDTCQDSSGGLPMYQSDQ-WHVVGIVSWGYCGGSPSTPGVYTKVSAYLNIWVNV- 279
  ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
b 234 FEEGKKDACKDSSGGLVCLVQSWLQAGVISWEGGCARQNRPGVIIRVTAHNNWIHRII 293
  ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Y 280 -----WKAELSRHHHHH 292
  :|||||

b 294 PKLQFQPARLGGQKSRHHHHH 315

```

RESULT 5
 S-09-386-642-11
 Sequence 11, Application US/09386642
 Patent No. 6420157
 GENERAL INFORMATION:
 APPLICANT: Darrow, Andrew
 APPLICANT: Qi, Jensen
 APPLICANT: Andrade-Gordon, Patricia
 TITLE OF INVENTION: Zymogen Activation System
 FILE REFERENCE: ORT-1028
 CURRENT APPLICATION NUMBER: US/09/386,642
 CURRENT FILING DATE: 1999-08-31
 NUMBER OF SEQ ID NOS: 60
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 11
 LENGTH: 328
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
 OTHER INFORMATION: with homo sapien serine protease catalytic domain
 S-09-386-642-11

```
Query Match          46.6%; Score 746; DB 4; Length 328;  
Best Local Similarity 49.7%; Pred. No. 4e-72;  
Matches 163; Conservative 35; Mismatches 94; Indels 36; Gaps 8;
```

y	1	MDSKGSSQSRLLLLLVSNLLLCQGVVSDYKDDDDVDAAALAAPFDDDKIVGGYALDV	60
b	1	MDSKGSSQSRLLLLLVSNLLLCQGVVSDYKDDDDVDAAALAAPFDDDKIVGGYALEA	60
y	61	DSWPWQVSIOYDQKHVCVCGSILDPHWLTAAHCFRKHTDFVNWKVRAGSKLGSPFLA-	119
b	61	GQWPWQVSIITVEGHVHCVCGSLVSEQWLSAHCFPSEHHKEAYEVKLGAHQLDYSYEDAK	120
y	120	VAKIIIEFNMPYKD--NDIALMKLPPLTFSGTVRPICLPFFDEELTPATPLWIIGW	176
b	121	VSTLKDIIPHPSYLQEGSQGDIALQLSRPITFSRYIRPICLPAANASFPNGLHCTVTGW	180
y	177	GFTKQNGGKMS-DILLQASVQVIDSTRCNA---DDAYOQE---VTEKMWCAGIPEGGVDT	229
b	181	GHVAPSVSLLTPKPQQLEVPILISRETNCNLINIDAKPEEHFHVEDMVCAGYVEGGKDA	240

```

QY      230 CQDGGGLMYQSD-QWHVVGVISWVGCGGPGSTPGVYTKVSAYLNWIY----- 277
          ||||||| : |::||| | | : ||||| | : | |
Db      241 CQDGGGPLSCPVEGLWYLTLGIVSWGDCAGARNRPGVYTCLASSYASWIQSVTELQPRVV 300

QY      278 -----NVWKAEI---SRHHHHH 292
          | : | | |||||
Db      301 PQTESQPDSNLCGSHLAFSSRRHHHHH 328


RESULT 6
US-09-386-629-8
; Sequence 8, Application US/09386629
; Patent No. 6426199
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L.
; APPLICANT: Qi, Jensen
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Identification and Characterization of the complementary
;   TITLE OF INVENTION: DNA encoding the novel human serine protease C-E
; FILE REFERENCE: ORT-1030
; CURRENT APPLICATION NUMBER: US/09/386,629
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: C-E catalytic
;   OTHER INFORMATION: domain fusion protien
US-09-386-629-8
```

RESULT 7
US-09-386-642-14
; Sequence 14, Application US/09386642
; Patent No. 6420157
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jenson
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Zymogen Activation System
; FILE REFERENCE: ORT-1028
; CURRENT APPLICATION NUMBER: US/09/386,642

CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 289
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
OTHER INFORMATION: with homo sapien serine protease catalytic domain
S-09-386-642-14

Query Match 42.0%; Score 672.5; DB 4; Length 289;
Best Local Similarity 48.5%; Pred. No. 2.8e-64;
Matches 148; Conservative 39; Mismatches 89; Indels 29; Gaps 11;

Y 1 MDSKSSQKSRLLLLLVSNLLLCQGVVSDYKDDDDVDAALAAPFDDDDKIVGGY-ALD 59
b 1 MDSKSSQKSRLLLLLVSNLLLCQGVVSDYKDDDDVDAALAAPFDDDDKIVGGY-ALD 59
Y 60 VDSWPWQVSIQYDKQHVCGGSILDPHWLTAACHCFR-----HTDVFVNWKVRAGSDK---- 111
b 61 KHSQPWQAALFEKTRLLCGATLIAFPWLLTAACHCLKPRYIVHLGQHNLQKEGCEQTRTA 120
Y 112 LGSFPLAVAKIIIEFNPMY-KD--NDIALMKLQFPLTFSGTVRPICLPFFDEELTPA 168
b 121 TESFPHPG-----FNNSLPNKDHRNDIMLVKMASPVSIWAVRPLTSL--SRCVTAG 170
Y 169 TPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMKAGIPEGGVD 228
b 171 TSCLISGWSTSSPOLRLPHTLRCANITIEHQKC--ENAYPGNITDTMVCASVQEGGKD 228
Y 229 TCQDSDGGPLMYQSDQWHVGVISWGYG-CGGPSTPGVYTKVSAYLNWIYNVWKAELSRH 287
b 229 SCQDSDGGPLVCNQSG--LQGIISWQDQPCAITRKPGVYTKVCKYVDWIQETMKN-SRH 284
Y 288 HHHH 292
b 285 HHHH 289

RESULT 8
S-09-386-642-13
Sequence 13, Application US/09386642
Patent No. 6420157
GENERAL INFORMATION:
APPLICANT: Darrow, Andrew
APPLICANT: Qi, Jensen
APPLICANT: Andrade-Gordon, Patricia
TITLE OF INVENTION: Zymogen Activation System
FILE REFERENCE: ORT-1028
CURRENT APPLICATION NUMBER: US/09/386,642
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 288
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
OTHER INFORMATION: with homo sapien serine protease catalytic domain
S-09-386-642-13

Query Match 40.3%; Score 645; DB 4; Length 288;
Best Local Similarity 47.9%; Pred. No. 2.6e-61;
Matches 146; Conservative 38; Mismatches 91; Indels 30; Gaps 11;

Y 1 MDSKSSQKSRLLLLLVSNLLLCQGVVSDYKDDDDVDAALAAPFDDDDKIVGGY-ALD 59
b 1 MDSKSSQKSRLLLLLVSNLLLCQGVVSDYKDDDDVDAALAAPFDDDDKIVGGY-ALD 59
Y 60 VDSWPWQVSIQYDKQHVCGGSILDPHWLTAACHCFR-KHTDVFVNWKVRAGSKL----GS 114

Db 61 PHSQPWQAALFQGOQLLCGVLVGNWVLTAAHCKPKYT-----VRLGDHSLQNKDGP 114
QY 115 FPSLAVAKIIIEFNPMY-----PKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPA 168
Db 115 EQEIPVVSIP--HPCYNSSDVEDHNHDLMLQLLDQASLGSVKVPISLA--DHCTQPG 169
QY 169 TPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMKAGIPEGGVD 228
Db 170 QKCTVSGWGTVTSPRENFPDTLNCAEVKIFPQKCC--EDAYPGQITDGMVCAGSSK-GAD 226
QY 229 TCQDSDGGPLMYQSDQWHVGVISWGYG-CGGPSTPGVYTKVSAYLNWIYNVWKAELSRH 287
Db 227 TCQDSDGGPLVCDG---ALQGITSWGSDDPCGRSDKPGVYTNICRYLDWIKKIIGSKGSRH 283
QY 288 HHHH 292
Db 284 HHHH 288

RESULT 9
US-09-386-642-53
Sequence 53, Application US/09386642
Patent No. 6420157
GENERAL INFORMATION:
APPLICANT: Darrow, Andrew
APPLICANT: Qi, Jensen
APPLICANT: Andrade-Gordon, Patricia
TITLE OF INVENTION: Zymogen Activation System
FILE REFERENCE: ORT-1028
CURRENT APPLICATION NUMBER: US/09/386,642
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 53
LENGTH: 306
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Fusion gene of
OTHER INFORMATION: human protease F in CPEK2 zymogen vector
US-09-386-642-53

Query Match 37.6%; Score 601.5; DB 4; Length 306;
Best Local Similarity 42.9%; Pred. No. 1.3e-56;
Matches 126; Conservative 46; Mismatches 85; Indels 37; Gaps 10;

QY 28 VSDYKDDDDVDAALAAPFDDDDKIVGGYALDVDSWPWQVSIQYDKQHVCGGSILDPHW 87
Db 21 VPDYKDD--DAAALAAPFDDDDKIVGGYALELGRWPWQGSRLWDSHVCGVSLLSHRWA 78
QY 88 LTAACHCFRKHTDVFN--WKVRAGSDKLGSPSLAVAK-----IIIEFNPMYPKON- 136
Db 79 LTAACHFETYSDLSDPSGWMVQFG--QLTSMPSFWSLQAYYNYFVSNYILSPRYLGNSP 136
QY 137 -DIALMKLQFPLTFSGTVRPICLPFFDEELTPATPLWIIGWFTKQNGKMSDILLQ-AS 194
Db 137 YDIALVKLSAPVYTYTHIQIPICLQASTFEFENRTDCWVTGWYIKEDALPSPHTLQEVQ 196
QY 195 VQVIDSTRCN---ADDAYQGEVTEKMMKAGIPEGGVDTCQDSDGGPL-MYQSDQWHVVG 250
Db 197 VAIINNSMCNHLFLKYFRKDI FGMVVCAGNAQGGKDACFGDGGGLACNKNGLWYQIGV 256
QY 251 VSWGYGCGGPSTPGVYTKVSAYLNWIYNV-----WKAELSRHHHHH 292
Db 257 VSWGVGCGRPNRPGVYTNISHHFEWIKLMAQSGMSQPDPSW-----SRHHHHH 306

RESULT 10
US-09-386-642-12
Sequence 12, Application US/09386642
Patent No. 6420157
GENERAL INFORMATION:

S-09-386-642-12

Query Match 32.9%; Score 526.5; DB 1; Length 418;
Best Local Similarity 41.8%; Pred. NO. 2.Se-48;
Matches 97; Conservative 49; Mismatches 79; Indels 7

b 302 PGSTAYVTGWGAQYAGHTVPE-LRQGVRIISNDVCNAPHSYNGAILSGMLCAGVPQGG 360
Y 227 VDTCCGDSGGPLMYQSDQ--WHVVGIVSWGCGGPGSTPGVYTKVSAYLNI 276
b 361 VDACCGDSGGPLVQEDSRRLWFIVGIVSWGDCGLPDKPGVYTRVTAYIDWI 412

earch completed: November 23, 2003, 07:53:47
ob time : 15.8611 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - protein search, using sw model

run on: November 23, 2003, 07:49:15 ; Search time 18.4759 Seconds
(without alignments)
2885.234 Million cell updates/sec

itle: US-09-607-745-9
erfect score: 1600
equence: 1 MDSKGSSQKSRLLLLLVVN.....LNWIYNVWKAELSRHHHHH 292

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 666188 seqs, 18259486 residues

otal number of hits satisfying chosen parameters: 666188

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*

7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*

10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*

11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*

13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*

14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*

15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*

16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*

17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1266	79.1	406	9	US-09-851-588-6
2	1266	79.1	423	11	US-09-776-191-72
3	1266	79.1	423	15	US-10-264-820-23
4	1266	79.1	423	15	US-10-254-289-2
5	1266	79.1	432	11	US-09-888-257A-7
6	1266	79.1	432	11	US-09-946-374-275
7	1266	79.1	432	12	US-10-015-387A-275
8	1266	79.1	432	12	US-10-063-735-112
9	1266	79.1	432	12	US-10-006-130A-275
10	1266	79.1	432	12	US-10-199-672-330
11	1266	79.1	432	12	US-10-006-172A-275
12	1266	79.1	432	12	US-10-187-749-330
13	1266	79.1	432	12	US-10-194-457-330
14	1266	79.1	432	12	US-10-184-642-330
15	1266	79.1	432	12	US-10-196-747-330

16	1266	79.1	432	12	US-10-015-392A-275	Sequence 275, App
17	1266	79.1	432	12	US-10-017-253A-275	Sequence 275, App
18	1266	79.1	432	12	US-10-173-689-330	Sequence 330, App
19	1266	79.1	432	12	US-10-173-690-330	Sequence 330, App
20	1266	79.1	432	12	US-10-173-691-330	Sequence 330, App
21	1266	79.1	432	12	US-10-173-692-330	Sequence 330, App
22	1266	79.1	432	12	US-10-173-694-330	Sequence 330, App
23	1266	79.1	432	12	US-10-173-698-330	Sequence 330, App
24	1266	79.1	432	12	US-10-173-699-330	Sequence 330, App
25	1266	79.1	432	12	US-10-173-707-330	Sequence 330, App
26	1266	79.1	432	12	US-10-174-569-330	Sequence 330, App
27	1266	79.1	432	12	US-10-174-583-330	Sequence 330, App
28	1266	79.1	432	12	US-10-174-587-330	Sequence 330, App
29	1266	79.1	432	12	US-10-174-589-330	Sequence 330, App
30	1266	79.1	432	12	US-10-174-591-330	Sequence 330, App
31	1266	79.1	432	12	US-10-175-736-330	Sequence 330, App
32	1266	79.1	432	12	US-10-175-742-330	Sequence 330, App
33	1266	79.1	432	12	US-10-175-744-330	Sequence 330, App
34	1266	79.1	432	12	US-10-175-745-330	Sequence 330, App
35	1266	79.1	432	12	US-10-175-748-330	Sequence 330, App
36	1266	79.1	432	12	US-10-175-751-330	Sequence 330, App
37	1266	79.1	432	12	US-10-175-754-330	Sequence 330, App
38	1266	79.1	432	12	US-10-176-480-330	Sequence 330, App
39	1266	79.1	432	12	US-10-176-489-330	Sequence 330, App
40	1266	79.1	432	12	US-10-176-754-330	Sequence 330, App
41	1266	79.1	432	12	US-10-176-755-330	Sequence 330, App
42	1266	79.1	432	12	US-10-176-759-330	Sequence 330, App
43	1266	79.1	432	12	US-10-176-920-330	Sequence 330, App
44	1266	79.1	432	12	US-10-176-922-330	Sequence 330, App
45	1266	79.1	432	12	US-10-176-924-330	Sequence 330, App

ALIGNMENTS

RESULT 1

US-09-851-588-6

; Sequence 6, Application US/09851588

; Patent No. US20020042067A1

; GENERAL INFORMATION:

; APPLICANT: Mack, David

; APPLICANT: Gish, Kurt C.

; APPLICANT: Wilson, Keith E.

; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND

; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS

; FILE REFERENCE: A-68829-1/DJB/JUD/AMS

; CURRENT APPLICATION NUMBER: US/09/851,588

; CURRENT FILING DATE: 2001-09-24

; PRIOR APPLICATION NUMBER: US 09/642,252

; PRIOR FILING DATE: 2000-08-17

; PRIOR APPLICATION NUMBER: US 09/656,002

; PRIOR FILING DATE: 2000-09-06

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 6

; LENGTH: 406

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-851-588-6

Query Match 79.1%; Score 1266; DB 9; Length 406;
Best Local Similarity 97.4%; Pred. No. 5.8e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY	51	KIVGYALDVDSWPWQVSIQYDKQHVCGSILDPHWLTAHCFRKHDTVFNWKVRAGSD	110
Db	173	RVYGGEEASVDSWPWQVSIQYDKQHVCGSILDPHWLTAHCFRKHDTVFNWKVRAGSD	232
QY	111	KLGSFPLAVAKIIIEFNPMYPKNDIALMKIQPLTFSGTVRPICLPFFDEELTPATP	170
Db	233	KLGSFPLAVAKIIIEFNPMYPKNDIALMKIQPLTFSGTVRPICLPFFDEELTPATP	292
QY	171	LWIGWGFTKQNGGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMCAGIPEGGVDTG	230

b 293 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGVDTC 352
y 231 QGDSGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVSAYLNWIYNVWKAE 284
b 353 QGDSGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVSAYLNWIYNVWKAE 406

RESULT 2
S-09-776-191-72
Sequence 72, Application US/09776191
Publication No. US20030119168A1
GENERAL INFORMATION:
APPLICANT: Edwin L. Madison
APPLICANT: Edgar O. Ong
APPLICANT: Jiumn-Chern Yeh
APPLICANT: Corvas International, Inc.
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
TITLE OF INVENTION: METHODS BASED THEREON
FILE REFERENCE: 24745-1607
CURRENT APPLICATION NUMBER: US/09/776,191
CURRENT FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/213,124
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 60/234,840
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 60/179,982
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: 60/183,542
PRIOR FILING DATE: 2000-02-18
PRIOR APPLICATION NUMBER: 09/657,968
PRIOR FILING DATE: 2000-02-08
NUMBER OF SEQ ID NOS: 72
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 72
LENGTH: 423
TYPE: PRT
ORGANISM: Homo sapien
S-09-776-191-72

Query Match 79.1%; Score 1266; DB 11; Length 423;
Best Local Similarity 97.4%; Pred. No. 6.1e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

y 51 KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWLVTAACHCFRKHDTDFVFNWKVRAGSD 110
b 190 RVVGGEASVDSWPQVSIQYDKQHVCGGSILDPHWLVTAACHCFRKHDTDFVFNWKVRAGSD 249
y 111 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVPICLPFFDEELTPATP 170
b 250 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVPICLPFFDEELTPATP 309
y 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGVDTC 230
b 310 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGVDTC 369
y 231 QGDSGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVSAYLNWIYNVWKAE 284
b 370 QGDSGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVSAYLNWIYNVWKAE 423

RESULT 3
US-10-264-820-23
Sequence 23, Application US/10264820
Publication No. US20030108926A1
GENERAL INFORMATION:
APPLICANT: Bos Biotechnology, Inc.
TITLE OF INVENTION: No. US20030108926A1 Methods of Diagnosing Colorectal Cancer,
Publication No. US20030108926A1
TITLE OF INVENTION: Compositions, and Methods of Screening for Colorectal
TITLE OF INVENTION: Cancer Modulators
FILE REFERENCE: 018501-006141US
CURRENT APPLICATION NUMBER: US/10/264,820

; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: US 09/268,866
; PRIOR FILING DATE: 1999-03-15
; PRIOR APPLICATION NUMBER: US 09/435,945
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: US 09/436,983
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: US 09/450,857
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: US 09/453,850
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human CGAS
US-10-264-820-23

Query Match 79.1%; Score 1266; DB 15; Length 423;
Best Local Similarity 97.4%; Pred. No. 6.1e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 51 KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWLVTAACHCFRKHDTDFVFNWKVRAGSD 110
Db 190 RVVGGEASVDSWPQVSIQYDKQHVCGGSILDPHWLVTAACHCFRKHDTDFVFNWKVRAGSD 249
QY 111 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVPICLPFFDEELTPATP 170
Db 250 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVPICLPFFDEELTPATP 309
QY 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGVDTC 230
Db 310 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGVDTC 369
QY 231 QGDSGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVSAYLNWIYNVWKAE 284
Db 370 QGDSGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVSAYLNWIYNVWKAE 423

RESULT 4
US-10-254-289-2
; Sequence 2, Application US/10254289
; Publication No. US20030118509A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-69108/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/10/254,289
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US/09/656,002
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens

S-10-254-289-2

Query Match 79.1%; Score 1266; DB 15; Length 423;
Best Local Similarity 97.4%; Pred. No. 6.1e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Y 51 KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWVLTAAHCFRKHDTDFNWKVRAGSD 110
:|||||
b 190 RVVGGEEASVDSWPQVSIQYDKQHVCGGSILDPHWVLTAAHCFRKHDTDFNWKVRAGSD 249
Y 111 KLGSPSLAVAKIIIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
b 250 KLGSPSLAVAKIIIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 309
Y 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQGEVTEKMCAGIPEGGVDTTC 230
b 310 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQGEVTEKMCAGIPEGGVDTTC 369
Y 231 QGDSGGLMYQSDQHVHVGVISWGYGCGGPSTPGVYTKVSAYLNWIYNVWKAEL 284
b 370 QGDSGGLMYQSDQHVHVGVISWGYGCGGPSTPGVYTKVSAYLNWIYNVWKAEL 423

RESULT 5

S-09-888-257A-7

Sequence 7, Application US/09888257A

Publication No. US20030060612A1

GENERAL INFORMATION:

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth J.

APPLICANT: Polakis, Paul

APPLICANT: Smith, Victoria

APPLICANT: Wood, William I.

APPLICANT: Wu, Thomas D.

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND

TITLE OF INVENTION: TREATMENT OF TUMOR

FILE REFERENCE: P5002R1

CURRENT APPLICATION NUMBER: US/09/888,257A

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/063,540

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: US 60/089,653

PRIOR FILING DATE: 1998-06-17

PRIOR APPLICATION NUMBER: US 60/099,792

PRIOR FILING DATE: 1998-09-10

PRIOR APPLICATION NUMBER: US 60/103,678

PRIOR FILING DATE: 1998-10-08

PRIOR APPLICATION NUMBER: US 60/235,451

PRIOR FILING DATE: 2000-09-26

PRIOR APPLICATION NUMBER: PCT/US99/12252

PRIOR FILING DATE: 1999-06-02

PRIOR APPLICATION NUMBER: PCT/US99/20111

PRIOR FILING DATE: 1999-09-01

PRIOR APPLICATION NUMBER: PCT/US00/04342

PRIOR FILING DATE: 2000-02-18

PRIOR APPLICATION NUMBER: PCT/US00/05841

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: PCT/US00/08439

PRIOR FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: PCT/US00/23328

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: PCT/US00/32678

PRIOR FILING DATE: 2000-12-01

PRIOR APPLICATION NUMBER: PCT/US01/06520

PRIOR FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: PCT/US01/06666

PRIOR FILING DATE: 2001-03-01

NUMBER OF SEQ ID NOS: 10

SEQ ID NO 7

LENGTH: 432

TYPE: PRT

ORGANISM: Homo Sapien

US-09-888-257A-7

Query Match 79.1%; Score 1266; DB 11; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 51 KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWVLTAAHCFRKHDTDFNWKVRAGSD 110
:|||||
Db 199 RVVGGEEASVDSWPQVSIQYDKQHVCGGSILDPHWVLTAAHCFRKHDTDFNWKVRAGSD 258
QY 111 KLGSPSLAVAKIIIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
Db 259 KLGSPSLAVAKIIIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 318
QY 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQGEVTEKMCAGIPEGGVDTTC 230
Db 319 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQGEVTEKMCAGIPEGGVDTTC 378
QY 231 QGDSGGLMYQSDQHVHVGVISWGYGCGGPSTPGVYTKVSAYLNWIYNVWKAEL 284
Db 379 QGDSGGLMYQSDQHVHVGVISWGYGCGGPSTPGVYTKVSAYLNWIYNVWKAEL 432

RESULT 6

US-09-946-374-275

Sequence 275, Application US/09946374

Publication No. US20030073129A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth J.

APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann

APPLICANT: Smith, Victoria

APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel

APPLICANT: Watanabe, Colin K.

APPLICANT: Williams, P. Mickey

APPLICANT: Wood, William I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

TITLE OF INVENTION: Acids Encoding the Same

FILE REFERENCE: P2830P1C1

CURRENT APPLICATION NUMBER: US/09/946,374

CURRENT FILING DATE: 2001-09-04

PRIOR APPLICATION NUMBER: 60/098716

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098723

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098749

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098750

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098803

PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/098821

PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/098843

PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/099536

PRIOR FILING DATE: 1998-09-09

PRIOR APPLICATION NUMBER: 60/099596

;	PRIOR APPLICATION NUMBER: 60/101477
;	PRIOR FILING DATE: 1998-09-23
;	PRIOR APPLICATION NUMBER: 60/101479
;	PRIOR FILING DATE: 1998-09-23
;	PRIOR APPLICATION NUMBER: 60/101738
;	PRIOR FILING DATE: 1998-09-24
;	PRIOR APPLICATION NUMBER: 60/101741
;	PRIOR FILING DATE: 1998-09-24
;	PRIOR APPLICATION NUMBER: 60/101743
;	PRIOR FILING DATE: 1998-09-24
;	PRIOR APPLICATION NUMBER: 60/101915
;	PRIOR FILING DATE: 1998-09-24
;	PRIOR APPLICATION NUMBER: 60/101916
;	PRIOR FILING DATE: 1998-09-24
;	PRIOR APPLICATION NUMBER: 60/102207
;	PRIOR FILING DATE: 1998-09-29
;	PRIOR APPLICATION NUMBER: 60/102240
;	PRIOR FILING DATE: 1998-09-29
;	PRIOR APPLICATION NUMBER: 60/102307
;	PRIOR FILING DATE: 1998-09-29
;	PRIOR APPLICATION NUMBER: 60/102330
;	PRIOR FILING DATE: 1998-09-29
;	PRIOR APPLICATION NUMBER: 60/102331
;	PRIOR FILING DATE: 1998-09-29
;	PRIOR APPLICATION NUMBER: 60/102484
;	PRIOR FILING DATE: 1998-09-30
;	PRIOR APPLICATION NUMBER: 60/102487
;	PRIOR FILING DATE: 1998-09-30
;	PRIOR APPLICATION NUMBER: 60/102570
;	PRIOR FILING DATE: 1998-09-30
;	PRIOR APPLICATION NUMBER: 60/102571
;	PRIOR FILING DATE: 1998-09-30
;	PRIOR APPLICATION NUMBER: 60/102684
;	PRIOR FILING DATE: 1998-10-01
;	PRIOR APPLICATION NUMBER: 60/102687
;	PRIOR FILING DATE: 1998-10-01
;	PRIOR APPLICATION NUMBER: 60/102965
;	PRIOR FILING DATE: 1998-10-02
;	PRIOR APPLICATION NUMBER: 60/103258
;	PRIOR FILING DATE: 1998-10-06
;	PRIOR APPLICATION NUMBER: 60/103314
;	PRIOR FILING DATE: 1998-10-07
;	PRIOR APPLICATION NUMBER: 60/103315
;	PRIOR FILING DATE: 1998-10-07
;	PRIOR APPLICATION NUMBER: 60/103328
;	PRIOR FILING DATE: 1998-10-07
;	PRIOR APPLICATION NUMBER: 60/103395
;	PRIOR FILING DATE: 1998-10-07
;	PRIOR APPLICATION NUMBER: 60/103396
;	PRIOR FILING DATE: 1998-10-07
;	PRIOR APPLICATION NUMBER: 60/103401
;	PRIOR FILING DATE: 1998-10-07
;	PRIOR APPLICATION NUMBER: 60/103449
;	PRIOR FILING DATE: 1998-10-06
;	PRIOR APPLICATION NUMBER: 60/103633
;	PRIOR FILING DATE: 1998-10-08
;	PRIOR APPLICATION NUMBER: 60/103678
;	PRIOR FILING DATE: 1998-10-08
;	PRIOR APPLICATION NUMBER: 60/103679
;	PRIOR FILING DATE: 1998-10-08
;	PRIOR APPLICATION NUMBER: 60/103711
;	PRIOR FILING DATE: 1998-10-08
;	PRIOR APPLICATION NUMBER: 60/104257
;	PRIOR FILING DATE: 1998-10-14
;	PRIOR APPLICATION NUMBER: 60/104987
;	PRIOR FILING DATE: 1998-10-20
;	PRIOR APPLICATION NUMBER: 60/105000
;	PRIOR FILING DATE: 1998-10-20
;	PRIOR APPLICATION NUMBER: 60/105002
;	PRIOR FILING DATE: 1998-10-20
;	PRIOR APPLICATION NUMBER: 60/105104
;	PRIOR FILING DATE: 1998-10-21
;	PRIOR APPLICATION NUMBER: 60/105169

PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105266
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105693
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105694
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105807

Query Match 79.1%; Score 1266; DB 11; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Y 51 KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHTDVFNWKVRAGSD 110
b 199 RVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHTDVFNWKVRAGSD 258
Y 111 KLGSPPSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 170
b 259 KLGSPPSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 318
Y 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGVDTC 230
b 319 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGVDTC 378
Y 231 QGDSGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 284
b 379 QGDSGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 432

ESULT 7
S-10-015-387A-275
Sequence 275, Application US/10015387A
Publication No. US20030135034A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan I.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2830PIC54
CURRENT APPLICATION NUMBER: US/10/015,387A
CURRENT FILING DATE: 2001-12-12
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 477
SEQ ID NO 275
LENGTH: 432
TYPE: PRT
ORGANISM: Homo sapiens
S-10-015-387A-275

Query Match 79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Y 51 KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHTDVFNWKVRAGSD 110
b 199 RVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHTDVFNWKVRAGSD 258
Y 111 KLGSPPSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 170
b 259 KLGSPPSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 318

QY 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGVDTC 230
Db 319 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGVDTC 378
QY 231 QGDSGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 284
Db 379 QGDSGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 432

RESULT 8
US-10-063-735-112
Sequence 112, Application US/10063735
Publication No. US2003013882A1
GENERAL INFORMATION:
APPLICANT: Eaton, Dan L.
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3230R1C1
CURRENT APPLICATION NUMBER: US/10/063,735
CURRENT FILING DATE: 2002-05-08
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 170
SEQ ID NO 112
LENGTH: 432
TYPE: PRT
ORGANISM: Homo Sapien
US-10-063-735-112

Query Match 79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 51 KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHTDVFNWKVRAGSD 110
Db 199 RVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHTDVFNWKVRAGSD 258
QY 111 KLGSPPSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 170
Db 259 KLGSPPSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 318
QY 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGVDTC 230
Db 319 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGVDTC 378
QY 231 QGDSGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 284
Db 379 QGDSGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 432

RESULT 9
US-10-006-130A-275
Sequence 275, Application US/10006130A
Publication No. US20030148375A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan I.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2830P1C7
CURRENT APPLICATION NUMBER: US/10/006,130A
CURRENT FILING DATE: 2002-03-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 477
SEQ ID NO 275
LENGTH: 432
TYPE: PRT
ORGANISM: Homo sapiens
S-10-006-130A-275

Query Match 79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Y 51 KIVGGYALDVSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHTDVFNWKVRAGSD 110
: : : : :
b 199 RVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHTDVFNWKVRAGSD 258
: : : : :
Y 111 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
: : : : :
b 259 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 318
: : : : :
Y 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQGEVTEKMMCAGIPEGGVDT 230
: : : : :
b 319 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQGEVTEKMMCAGIPEGGVDT 378
: : : : :
Y 231 QGDSGGPLMYQSDQHVVGIVSWGCGGPSTPGVYTKVSAYLNWIYNVWKAEL 284
: : : : :
b 379 QGDSGGPLMYQSDQHVVGIVSWGCGGPSTPGVYTKVSAYLNWIYNVWKAEL 432
: : : : :

RESULT 10
S-10-199-672-330
Sequence 330, Application US/10199672
Publication No. US20030148442A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/199,672
CURRENT FILING DATE: 2002-07-18
Prior Application Number: US/10/052,586
Prior Filing Date: 2002-01-15
Prior Application Number: 60/059263
Prior Filing Date: 1997-09-18
Prior Application Number: 60/059266
Prior Filing Date: 1997-09-18
Prior Application Number: 60/062250
Prior Filing Date: 1997-10-17
Prior Application Number: 60/063120
Prior Filing Date: 1997-10-24
Prior Application Number: 60/063121
Prior Filing Date: 1997-10-24
Prior Application Number: 60/063486
Prior Filing Date: 1997-10-21

; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-199-672-330

Query Match 79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 51 KIVGGYALDVSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHTDVFNWKVRAGSD 110
: : : : :
Db 199 RVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHTDVFNWKVRAGSD 258
: : : : :
QY 111 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
: : : : :
Db 259 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 318
: : : : :
QY 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQGEVTEKMMCAGIPEGGVDT 230
: : : : :
Db 319 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQGEVTEKMMCAGIPEGGVDT 378
: : : : :
QY 231 QGDSGGPLMYQSDQHVVGIVSWGCGGPSTPGVYTKVSAYLNWIYNVWKAEL 284
: : : : :
Db 379 QGDSGGPLMYQSDQHVVGIVSWGCGGPSTPGVYTKVSAYLNWIYNVWKAEL 432
: : : : :

RESULT 11
US-10-006-172A-275
Sequence 275, Application US/10006172A
Publication No. US20030153000A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan I.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2830P1C11
CURRENT APPLICATION NUMBER: US/10/006,172A
CURRENT FILING DATE: 2002-03-19
Prior Application Number: 60/098716
Prior Filing Date: 1998-09-01
Prior Application Number: 60/098723
Prior Filing Date: 1998-09-01
Prior Application Number: 60/098749
Prior Filing Date: 1998-09-01
Prior Application Number: 60/098750
Prior Filing Date: 1998-09-01
Prior Application Number: 60/098803
Prior Filing Date: 1998-09-02
Prior Application Number: 60/098821
Prior Filing Date: 1998-09-02
Prior Application Number: 60/098843
Prior Filing Date: 1998-09-02

PRIOR APPLICATION NUMBER: 60/105104
PRIOR FILING DATE: 1998-10-21
PRIOR APPLICATION NUMBER: 60/105169
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105266
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105693
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105694
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105807
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/105881
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/105882
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/106023
PRIOR FILING DATE: 1998-10-28

Query Match 79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

y 51 KIVGGYALDVSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHDTDFNWKVRAGSD 110
: : : : :
b 199 RVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHDTDFNWKVRAGSD 258
: : : : :
y 111 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
: : : : :
b 259 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 318
: : : : :
y 171 LMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQGEVTEKMMCAGIPEGGVDTTC 230
: : : : :
b 319 LMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQGEVTEKMMCAGIPEGGVDTTC 378
: : : : :
y 231 QGDSGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVSAYLNWIYNVWKAEL 284
: : : : :
b 379 QGDSGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVSAYLNWIYNVWKAEL 432
: : : : :

RESULT 12
US-10-187-749-330
Sequence 330, Application US/10187749
Publication No. US20030153036A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/187,749
CURRENT FILING DATE: 2002-07-01
PRIOR APPLICATION NUMBER: US/10/052,586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 330
LENGTH: 432
TYPE: PRT
ORGANISM: Homo Sapien
US-10-187-749-330

Query Match 79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 51 KIVGGYALDVSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHDTDFNWKVRAGSD 110
: : : : :
Db 199 RVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHDTDFNWKVRAGSD 258
: : : : :
QY 111 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
: : : : :
Db 259 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 318
: : : : :
QY 171 LMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQGEVTEKMMCAGIPEGGVDTTC 230
: : : : :
Db 319 LMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQGEVTEKMMCAGIPEGGVDTTC 378
: : : : :
QY 231 QGDSGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVSAYLNWIYNVWKAEL 284
: : : : :
Db 379 QGDSGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVSAYLNWIYNVWKAEL 432
: : : : :

RESULT 13
US-10-194-457-330
Sequence 330, Application US/10194457
Publication No. US20030153037A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C296
CURRENT APPLICATION NUMBER: US/10/194,457
CURRENT FILING DATE: 2002-07-11
PRIOR APPLICATION NUMBER: 10/052586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540

PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 330
LENGTH: 432
TYPE: PRT
ORGANISM: Homo Sapien
S-10-194-457-330

Query Match 79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Y 51 KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWLTAAHCFRKHTDVFNWKVRAGSD 110
: : : : :
b 199 RVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTAAHCFRKHTDVFNWKVRAGSD 258
Y 111 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
: : : : :
b 259 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 318
Y 171 LWIIGWGFTKQGGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMMCAGIPEGGVDT 230
: : : : :
b 319 LWIIGWGFTKQGGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMMCAGIPEGGVDT 378
Y 231 QGDSGGPLMYQSDQWHVVGIVSWGCGGPSTPGVYTKVSAYLNWIYNVWKAE 284
: : : : :
b 379 QGDSGGPLMYQSDQWHVVGIVSWGCGGPSTPGVYTKVSAYLNWIYNVWKAE 432

RESULT 14
S-10-184-642-330
Sequence 330, Application US/10184642
Publication No. US20030157635A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C194
CURRENT APPLICATION NUMBER: US/10/184,642
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 330
LENGTH: 432
TYPE: PRT
ORGANISM: Homo Sapien
S-10-184-642-330

Query Match 79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Y 51 KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWLTAAHCFRKHTDVFNWKVRAGSD 110
: : : : :
b 199 RVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTAAHCFRKHTDVFNWKVRAGSD 258
Y 111 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
: : : : :

Db 259 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 318
QY 171 LWIIGWGFTKQGGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMMCAGIPEGGVDT 230
: : : : :
Db 319 LWIIGWGFTKQGGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMMCAGIPEGGVDT 378
QY 231 QGDSGGPLMYQSDQWHVVGIVSWGCGGPSTPGVYTKVSAYLNWIYNVWKAE 284
: : : : :
Db 379 QGDSGGPLMYQSDQWHVVGIVSWGCGGPSTPGVYTKVSAYLNWIYNVWKAE 432
RESULT 15
US-10-196-747-330
; Sequence 330, Application US/10196747
; Publication No. US20030162250A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C346
; CURRENT APPLICATION NUMBER: US/10/196,747
; CURRENT FILING DATE: 2002-07-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-196-747-330

Query Match 79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 51 KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWLTAAHCFRKHTDVFNWKVRAGSD 110
: : : : :
Db 199 RVVGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTAAHCFRKHTDVFNWKVRAGSD 258
QY 111 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
: : : : :
Db 259 KLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 318
QY 171 LWIIGWGFTKQGGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMMCAGIPEGGVDT 230
: : : : :
Db 319 LWIIGWGFTKQGGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMMCAGIPEGGVDT 378
QY 231 QGDSGGPLMYQSDQWHVVGIVSWGCGGPSTPGVYTKVSAYLNWIYNVWKAE 284
: : : : :
Db 379 QGDSGGPLMYQSDQWHVVGIVSWGCGGPSTPGVYTKVSAYLNWIYNVWKAE 432

Search completed: November 23, 2003, 07:54:48
Job time : 19.4759 secs

crosin (EC 3.4.21.10) precursor form 3 - mouse
;Species: Mus musculus (house mouse)
;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Jun-2000
;Accession: JX0172; JX0138
;Watanabe, K.; Baba, T.; Kashiwabara, S.; Okamoto, A.; Arai, Y.
;Biochem. 109, 828-833, 1991
;Title: Structure and organization of the mouse acrosin gene.
;Reference number: JX0172; MUID:92041732; PMID:1939002
;Accession: JX0172
;Molecule type: DNA
;Residues: 1-436 <WAT>
;Cross-references: GB:S66245; NID:G238706; PIDN:AAB20293.1; PID:G238707
;Kashiwabara, S.; Baba, T.; Takada, M.; Watanabe, K.; Yano, Y.; Arai, Y.
;Biochem. 108, 785-791, 1990
;Title: Primary structure of mouse proacrosin deduced from the cDNA sequence and its ge
;Reference number: JX0138; MUID:91185335; PMID:2127931
;Accession: JX0138
;Molecule type: mRNA
;Residues: 4-436 <KAS>
;Cross-references: GB:D00754; NID:G220322; PIDN:BAA00651.1; PID:G220323
;Comment: Acrosin is an acrosomal protease that plays an important role in the initial
-binding activity.
;Genetics:
;Introns: 26/2; 95/2; 190/1; 238/3
;Superfamily: acrosin; trypsin homology
;Keywords: glycoprotein; hydrolase; serine proteinase; sperm
;1-19/Domain: signal sequence #status predicted <SIG>
;20-42/Product: acrosin light chain #status predicted <ALC>
;43-321/Product: acrosin heavy chain #status predicted <AHC>
;43-286/Domain: trypsin homology <TRY>
;22,211/Binding site: carbohydrate (Asn) (covalent) #status predicted
;25-155,29-162,74-90,178-247,210-226,237-267/Disulfide bonds: #status predicted
;89,143,241/Active site: His, Asp, Ser #status predicted
;235/Binding site: substrate (Asp) #status predicted

Query Match 31.2%; Score 499; DB 2; Length 436;
Best Local Similarity 39.0%; Pred. No. 5.9e-37;
Matches 112; Conservative 43; Mismatches 104; Indels 28; Gaps 9;

Y 14 LLLVSNLLLCQGVVSDYKDDDDVDAALAAPFDDDD---KIVGGYALDVDSWPWQVSI 69
b 4 MLPTVAVLVAVSVA--KDNITCDGPC-GLRFRQNSQAGTRIVSGQSAQLGAWPMVSL 60

Y 70 QY-----DKQHVCGGSILDPHWLTAHCFRKHDTDFENWK-----VRAGSDKLGSP 116
b 61 QIFTSNRSRRYHACGGSLLNSHWLTAHCFDNKKKVDYDWRLVFGAQEIYGRNKPVEP 120

Y 117 SLA--VAKIIIEFNPMYPKNDIALMKLPPLTFSGTVRPICLPFFDEELTPATP--LW 172
b 121 QQERYVQKIVIEHKYNVVTGNDIALKTPPVTTCGNFIGPCCLPHF-KAGPPQIPHTCY 179

Y 173 IIGWFTKQNGKMSDILLQASVQVIDSTFCNADDAAYQGEVTERKNCAGIPEGGVDTCCG 232
b 180 VTGWGYIKEKAPRSPVLMPEARVDLIDLCLNSTQWYNGRVTSTNVCAGYPEGKIDTCG 239

Y 233 DSGGPLYMQSD---QWHVVGVISWGYGCGGPSTPGVYTKVSAVLNWI 276
b 240 DSGGPLMCRDNDVSPFVVVGITSGVGCARAKRPGVYTATWDYLDWI 286

RESULT 3
37344
crosin (EC 3.4.21.10) precursor form 1 - mouse
;Species: Mus musculus (house mouse)
;Date: 19-Mar-1993 #sequence_revision 19-Mar-1993 #text_change 22-Jun-1999
;Accession: A37344
;Klemm, U.; Maier, W.M.; Tsacousidou, S.; Adham, I.M.; Willison, K.; Engel, W.
;Differentiation 42, 160-166, 1990
;Title: Mouse preproacrosin: cDNA sequence, primary structure and postmeiotic expressio
;Reference number: A37344; MUID:90255839; PMID:2111255
;Accession: A37344
;Status: preliminary
;Molecule type: mRNA

A;Residues: 1-418 <KLE>
A;Cross-references: GB:X52466; NID:G49857; PIDN:CAA36704.1; PID:G49858
C;Superfamily: acrosin; trypsin homology
C;Keywords: glycoprotein; hydrolase; serine proteinase
F;39-283/Domain: trypsin homology <TRY>
F;18,208/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;21-151/Disulfide bonds: #status predicted
F;25-158/Disulfide bonds: #status predicted
F;70-86/Disulfide bonds: #status predicted
F;85,139,238/Active site: His, Asp, Ser #status predicted
F;175-244/Disulfide bonds: #status predicted
F;207-223/Disulfide bonds: #status predicted
F;234-264/Disulfide bonds: #status predicted

Query Match 30.8%; Score 493; DB 2; Length 418;
Best Local Similarity 41.5%; Pred. No. 2e-36;
Matches 102; Conservative 35; Mismatches 89; Indels 20; Gaps 5;

QY 51 KIVGGYALDVDSWPWQVSIQY-----DKQHVCGGSILDPHWLTAHCFRKHDTDFENWK 104
Db 38 RIVSGQSAHVGAWPWMVSLQIFTSNRSRRYHACGGSLLNSHWLTAHCFDNKKKVDYDWR 97

QY 105 -----VRAGSDKLGSPF--SLAVAKIIIEFNPMYPKNDIALMKLPPLTFSGTVRP 155
Db 98 LVFGAQEIYGRNKPVEKPEQEEYVQKIVIEHKYNVVTGNDIALKTPPVTTCGNFIGP 157

QY 156 ICLPFFDEELTPATP--LWIIWGFTKQNGKMSDILLQASVQVIDSTFCNADDAAYQGEV 213
Db 158 CCLPHFKAGPPRKIPHTCYVTGWGYIKREAPRSPVLMPEARVDLIDLCLNSTQWYNGRV 217

QY 214 TEKMMCAGIPEGGVDTCCQDSGGPLMYQSD---QWHVVGVISWGYGCGGPSTPGVYTKVS 270
Db 218 TSTNVCAGYPEGKIDTCQDSGGPLMCRDNRARQPFVVVGITSGVGCARAKRPGVYTATW 277

QY 271 AYLNWI 276
Db 278 DYLDWI 283

RESULT 4
KQMSPL
plasma kallikrein (EC 3.4.21.34) precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
C;Accession: A36557
R;Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochemont
A Cell Biol. 9, 737-748, 1990
A;Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and compariso
A;Reference number: A36557; MUID:91090844; PMID:2264928
A;Accession: A36557
A;Molecule type: mRNA
A;Residues: 1-638 <SEI>
A;Cross-references: GB:M58588; NID:G200358; PIDN:AAA63393.1; PID:G200359
A;Note: part of this sequence, including the amino ends of both the heavy and light chain
C;Comment: This protein, synthesized in the liver, circulates as a noncovalent complex wi
C;Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a liq
are linked by one or more disulfide bonds.
C;Superfamily: coagulation factor XI; trypsin homology
C;Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; inflan
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-390/Product: plasma kallikrein heavy chain #status experimental <HCH>
F;20-109/Domain: apple repeat <AP1>
F;110-199/Domain: apple repeat <AP2>
F;200-289/Domain: apple repeat <AP3>
F;291-380/Domain: apple repeat <AP4>
F;391-638/Product: plasma kallikrein light chain #status experimental <LCH>
F;391-621/Domain: trypsin homology <TRY>
F;21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-
F;127,215,308,396,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 30.8%; Score 492; DB 1; Length 638;
Best Local Similarity 43.5%; Pred. No. 4e-36;


```
;391-621/Domain: trypsin homology <TRY>
;21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322
;127,215,308,453,459,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
;396/Binding site: carbohydrate (Asn) (covalent) #status experimental
;434,483,578/Active site: His, Asp, Ser #status predicted

Query Match          30.6%; Score 489; DB 1; Length 638;
Best Local Similarity 42.2%; Pred. No. 7.4e-36;
Matches 100; Conservative 45; Mismatches 76; Indels 16; Gaps 8;

Y 51 KIVGGYALDVDSWPWQVSIQ---YDKQHVCGGSILDPHWVLTAAHCFR--KHDTDFVFNWKV 105
b 390 RIVGGTNSSSLGEWPWQVSLQVQLVSNQHMCGGSIIGRQWILTAAHCFDGIYPDV--WRI 447

Y 106 RAG-----SDKLGSPFLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFF 161
b 448 YGGILNLSEITNKTPTSSIKELIIHQYKXMSSEGYDIALIKLQTLNVTETFPQKPICLPSK 507

Y 162 DEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRONADDAAYQGEV-TEKMMCA 220
b 508 ADTNTIYTNWCVTGWTGER-GETQNILQKATIPLPVNEEC--QKKYRDYVITKQMICA 564

Y 221 GIPEGGVDTCCQDSGGPLMYQ-SDQHVHVGVISWGYCGGPGSTPGVYTKVSAYLNWI 276
b 565 GYKEGGIDACKDGGSGGPLVCKHSGRWQLVGTSGWEGECARKEQPGVYTKVAEYIDWI 621

RESULT 7
33777
epsin (EC 3.4.21.-) - rat
;Species: Rattus norvegicus (Norway rat)
;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
;Accession: S33777; S32013
;Farley, D.; Raymond, F.; Nick, H.
iochim. Biophys. Acta 1173, 350-352, 1993
;Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.
;Reference number: S33777; MUID:93305733; PMID:8318546
;Accession: S33777
;Status: preliminary
;Molecule type: mRNA
;Residues: 1-416 <FAR>
;Cross-references: EMBL:X70900; NID:g57928; PIDN:CAA50256.1; PID:g57929
;Superfamily: hepsin; trypsin homology
;Keywords: hydrolase; liver; serine proteinase; transmembrane protein
;22-44/Domain: transmembrane #status predicted <TMN>
;162-399/Domain: trypsin homology <TRY>
;187-203,290-358,321-337,348-380/Disulfide bonds: #status predicted
;202,256,352/Active site: His, Asp, Ser #status predicted

Query Match          30.0%; Score 479.5; DB 1; Length 416;
Best Local Similarity 39.8%; Pred. No. 3.2e-35;
Matches 98; Conservative 38; Mismatches 95; Indels 15; Gaps 4;

Y 50 DKIVGGYALDVDSWPWQVSIQYDKQHVCGGSILDPHWVLTAAHCFRKHDTDFVN-WKYRAG 108
b 160 DRIVGGQDSSLGRWPWQVSLRYDGTGLCGSLLSGDWVLTAAHCFERNRVLRSRWFAG 219

Y 109 SDKLGSPFLAVAKIIIEFNPMYP-----KNDIALMKLQFPLTFSGTVRPICLPFF 160
b 220 AVARTSPHAVQLGVQAVIYHGYLPFRDPTIDENSNDIALVELSSSLPLTEYIQPVCLPA 279

Y 161 FDEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRONADDAAYQGEVTERMMCA 220
b 280 AGQALVDGKVCVTGNGNT-QFYGQQAQVQLQEARVPIISNEVCNSPDFYGNQIKPMFCA 338

Y 221 GIPEGGVDTCCQDSGGPLMYQ-----SDQHVHVGVISWGYCGGPGSTPGVYTKVSAYLNW 275
b 339 GYPEGGIDACQDGGGHFVCEDRISGTSRWRLCGIVSWGTCALARKPGVYTKVIDPREW 398

Y 276 IYNVWK 281
b 399 IFQAIAK 404
```

```
RESULT 8
KQHUP
plasma kallikrein (EC 3.4.21.34) precursor - human
N;Alternate names: kininogenin; plasma prekallikrein
C;Species: Homo sapiens (man)
C;Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999
C;Accession: A00921; A37939
R;Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
Biochemistry 25, 2410-2417, 1986
A;Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four te
A;Reference number: A00921; MUID:86243359; PMID:3521732
A;Accession: A00921
A;Molecule type: mRNA
A;Residues: 1-638 <CHU>
A;Cross-references: GB:M13143; NID:g190262; PIDN:AAA60153.1; PID:g190263
R;McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2050-2056, 1991
A;Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of f
A;Reference number: A37939; MUID:91152016; PMID:1998666
A;Accession: A37939
A;Molecule type: protein
A;Residues: 20-27;40-46,'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80;103-113;131-140;141-1
;260-283,'X',285;287-291,'X',293-295;314-317,'X',319-320;321-324,'X',329-333;334-339,'X',
525;538-551;562,'X',564-567;573,'X',575-576;578-583,'X',585;592-604 <MCM>
C;Comment: This protein, synthesized in the liver, circulates as a noncovalent complex wj
C;Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a lig
are linked by one or more disulfide bonds.
C;Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal re
inogen and may also play a role in the renin-angiotensin system by converting prorenin ir
C;Genetics:
A;Gene: GDB:KLK3
A;Cross-references: GDB:127575; OMIM:229000
A;Map position: 4q35-4q35
C;Superfamily: coagulation factor XI; trypsin homology
C;Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; inflan
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-638/Product: plasma kallikrein #status predicted <MAT>
F;20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>
F;20-109/Domain: apple repeat <AP1>
F;110-199/Domain: apple repeat <AP2>
F;200-289/Domain: apple repeat <AP3>
F;291-380/Domain: apple repeat <AP4>
F;391-638/Domain: plasma kallikrein light chain #status predicted <LCH>
F;391-621/Domain: trypsin homology <TRY>
F;21-104;47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,322-328,383-
F;127,308,396,453,494/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;318-347,340-345/Disulfide bonds: #status predicted
F;390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted
F;434,483,578/Active site: His, Asp, Ser #status predicted

Query Match          29.9%; Score 479; DB 1; Length 638;
Best Local Similarity 41.8%; Pred. No. 5.8e-35;
Matches 99; Conservative 46; Mismatches 76; Indels 16; Gaps 8;

QY 51 KIVGGYALDVDSWPWQVSIQYD---KQHVCGGSILDPHWVLTAAHCF--RKHTDFVFNWKV 105
Db 390 RIVGGTNSSSGEWPWQVSLQVQLTAQRHLCGGSLIGHQWVLTAAHCFDGLPLQDV--WRI 447

QY 106 RAG-----SDKLGSPFLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFF 161
Db 448 YSGILNLSDITKDTPTFSQIKELIIHQYKXMSSEGNHDIALIKLQAPLNYTETFPQKPICLPSK 507

QY 162 DEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRONADDAAYQGEV-TEKMMCA 220
Db 508 GDTSTIYTNWCVTGWFSGKEK-GEIQNILQKVNPLVTVNEEC--QKRYQDYKITQRMVCA 564

QY 221 GIPEGGVDTCCQDSGGPLMYQ-SDQHVHVGVISWGYCGGPGSTPGVYTKVSAYLNWI 276
Db 565 GYKEGGKACKDGGSGGPLVCKHNGMWRLVGTSGWEGECARREQPGVYTKVAEYMDWI 621

RESULT 9
```


olyprotein - African clawed frog
;Species: Xenopus laevis (African clawed frog)
;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 03-Feb-2003
;Accession: T30337
;Yang, J.C.; Lindsay, L.L.; Hedrick, J.L.
ubmitted to the EMBL Data Library, March 1998
;Description: cDNA cloning of ovochymase, a chymotrypsin-like protease released from X
;Reference number: Z20829
;Accession: T30337
;Status: preliminary; translated from GB/EMBL/DBBJ
;Molecule type: mRNA
;Residues: 1-1524 <YAN>
;Cross-references: EMBL:U81290; NID:G2981640; PID:G2981641; PIDN:AAC24717.1
;Superfamily: tryosin related polypeptide; trypsin homology

Query Match 29.4%; Score 470; DB 2; Length 1524;
Best Local Similarity 43.0%; Pred. No. 1.1e-33;
Matches 101; Conservative 37; Mismatches 87; Indels 10; Gaps 7;

Y 51 KIVGGYALDVDSWPWQVSIQYDKQHVCGGSILDPHWLTAACHFRKHTDVFNNWKVRAGS- 109
b 583 RIVGEEASPNSWPWQVQIFFLRTFHCAGALISPPQWILTAACHCIRAAEPSY-WTVIAGDH 641
Y 110 DKLGSPFSLAVAKIIIEFNPMYPK---DNDIALMKLQFPLTFSGFVRPICLPFFDEELT 166
b 642 NRMLNESTEQIRNIKTIRIHDNYSNTYDNDIALLYLEPLDLNDFVRPCLPEPEVELT 701
Y 167 PATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQGEVTEKMMCAPIEG- 225
b 702 PASVCVVTGWNFAEDGQPALG-LQQLQLPLDSIIICNT-SYISSELTDHMLCAGFPSSK 759
Y 226 GVDTCQDGGPLMYQS--DQWVVGVISWYCGGPGSTPGVYTKVSAYLNWIYN 278
b 760 EKDACQDGGPLVCQNEKEQFSIYGLVSWGEGGRVSKPGVYTKVRLFFFTWION 814

ESULT 14
35340
trypsin (EC 3.4.21.4) 2 precursor - African malaria mosquito
;Species: Anopheles gambiae (African malaria mosquito)
;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 22-Jun-1999
;Accession: S35340; S40008; S35413
;Mueller, H.M.; Crampton, J.M.; della Torre, A.; Sindén, R.; Crisanti, A.
MBO J. 12, 2891-2900, 1993
;Title: Members of a trypsin gene family in Anopheles gambiae are induced in the gut by
;Reference number: S35339; MUID:93327779; PMID:8335004
;Accession: S35340
;Molecule type: DNA
;Residues: 1-277 <MUE>
;Cross-references: EMBL:Z18890; NID:G312023; PIDN:CAA79328.1; PID:G312024
;Experimental source: strain Suakoko
;Mueller, H.
ubmitted to the EMBL Data Library, June 1993
;Reference number: S40003
;Accession: S40008
;Molecule type: DNA
;Residues: 1-277 <MU2>
;Superfamily: trypsin; trypsin homology
;Keywords: hydrolase; protein digestion; serine proteinase; zymogen
;1-19/Domain: signal sequence #status predicted <SIG>
;20-50/Domain: activation peptide #status predicted <APT>
;51-277/Product: trypsin 2 #status predicted <MAT>
;51-271/Domain: trypsin homology <TRY>
;91,136,232/Active site: His, Asp, Ser #status predicted

Query Match 29.3%; Score 469; DB 2; Length 277;
Best Local Similarity 37.6%; Pred. No. 1.7e-34;
Matches 111; Conservative 46; Mismatches 88; Indels 50; Gaps 10;

Y 7 SQSRLLLLLVSNLLCQGVVSDYKDDDDVDAALAAPF----- 46
b 2 SNKIAILLVAVVAVAC-----AQAPSGRHLVHPLLPFLPRLHRD 45

QY 47 DDDDKIVGGYALDVDSWPWQVSIQYDKQHVCGGSILDPHWLTAACHFRKHTDVFNNWKVR 106
Db 46 SNGHRVVGGFQIDVSDAPYQVSLQYFNSHRCGGSVLDNKKWLTAAHC-TQGLDPSSLA VR 104
QY 107 AGSDKLGSPFSLAVAKIIIEFNPMYPK---DNDIALMKLQFPLTFSGFVRPICLPFFDE 163
Db 105 LGSSEHATGTL-VGVLRTVE-HPQYDGNITIDFSLMELETETLTFSDLVQVPELPEHEE 162
QY 164 ELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQ--GEVTEKMMCA 221
Db 163 PVEPGTMAVTSWGNT-QSAVESSEDFLRAANVPTVSHEDCS--DAYMWFGEITDRMLCAG 219
QY 222 IPEGGVDTCQDGGPLMYQSDQWVVGVISWYCGGPGSTPGVYTKVSAYLNWI 276
Db 220 YQGGKDACQDGGPLVADG---KLVGVVSWGYGCAQPGYVGRVASVRDWW 271

RESULT 15
B25852
trypsin (EC 3.4.21.4) II precursor [validated] - human
N;Alternate names: trypsin 2; trypsin, anionic; trypsinogen II
C;Species: Homo sapiens (man)
C;Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 08-Dec-2000
C;Accession: B25852; A61066; B43988
R;Emi, M.; Nakamura, Y.; Ogawa, M.; Yamamoto, T.; Nishide, T.; Mori, T.; Matsubara, K.
Gene 41, 305-310, 1986
A;Title: Cloning, characterization and nucleotide sequences of two cDNAs encoding human I
A;Reference number: A91544; MUID:86221712; PMID:3011602
A;Accession: B25852
A;Molecule type: mRNA
A;Residues: 1-247 <EMI>
A;Cross-references: GB:M27602; NID:G521217; PIDN:AAA61232.1; PID:G521218
R;Kimland, M.; Russick, C.; Marks, W.H.; Borgstroem, A.
Clin. Chim. Acta 184, 31-46, 1989
A;Title: Immunoreactive anionic and cationic trypsin in human serum.
A;Reference number: A61066; MUID:90091010; PMID:2598466
A;Accession: A61066
A;Molecule type: protein
A;Residues: 16-39,'X',41-42,'XXX',47-49 <KIM>
R;Koivunen, E.; Huhtala, M.L.; Stenman, U.H.
J. Biol. Chem. 264, 14095-14099, 1989
A;Title: Human ovarian tumor-associated trypsin. Its purification and characterization f
A;Reference number: A43988; MUID:89340515; PMID:2503510
A;Accession: B43988
A;Molecule type: protein
A;Residues: 16-49 <KOI>
A;Experimental source: mucinous ovarian tumor cyst fluid
C;Genetics:
A;Gene: GDB:PRSS2; TRY2
A;Cross-references: GDB:335289; OMIM:601564
A;Map position: 7q35-7q35
A;Note: the human genome contains at least ten trypsin genes or pseudogenes, at least tw
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-246/Product: trypsinogen II #status experimental <ZYM>
F;16-23/Domain: activation peptide #status experimental <APT>
F;24-246/Product: trypsin II #status predicted <ENZ>
F;24-239/Domain: trypsin homology <TRY>
F;30-160,48-64,171-185,196-220/Disulfide bonds: #status predicted
F;63,107,200/Active site: His, Asp, Ser #status predicted
F;75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 29.0%; Score 464.5; DB 1; Length 247;
Best Local Similarity 40.9%; Pred. No. 3.8e-34;
Matches 110; Conservative 32; Mismatches 84; Indels 43; Gaps 10;

QY 20 NLLCQGVSDYKDDDDVDAALAAPFDDDDKIVGGYALDVDSWPWQVSIQYDKQHVCGG 79
Db 2 NELLILTLTFV-----AAVAAPFDDDDKIVGGYICEENSVPYQVSLN-SGYHFCGG 50
QY 80 SILDPHWLTAACHFRKHTDV----FNWKVRAGSDKLGSPFSLAVAKIIIEFNPMYPK- 134

b 51 SLISEQWVVSAGHCYKSRIOVRLGEHNIEVLEGNQF-----INAAKII-----RHPKY 99
Y 135 -----DNDIALMKLQFPLTFSGTVRPICLPFFDELTTPA--TPLWIIIGWGFTKQNGGKMS 187
b 100 NSRTLNDNDILLIKLSSPAVINSRVSAISLP----TAPPAAGTESLISGWGNTLSSGADYP 155
Y 188 DILLQASVQVIDSTRCNADDAYQGEVTEKMMCAGIPEGGVDTCQGDSGGPLMYQSDQWHV 247
b 156 DELQCLDAPVLSQAECEA--SYPGKITNNMFCVGFLEGGKDSQCGDSGGFVVSNGE---L 210
Y 248 VGIVSWGYGCGGPSTPGVYTKVSAYLNWI 276
b 211 QGIVSWGYGCAQKXNRPVYTKVYNYVDWI 239

earch completed: November 23, 2003, 07:52:54
ob time : 20.2792 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - protein search, using sw model

un on: November 23, 2003, 07:39:05 ; Search time 11.2462 Seconds
(without alignments)
1221.015 Million cell updates/sec

itle: US-09-607-745-9

erfect score: 1600

equence: 1 MDSKGSQKSRLLLLLVSN.....LNWIYVWKALSRHHHHH 292

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

earched: 127863 seqs, 47026705 residues

otal number of hits satisfying chosen parameters: 127863

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1266	79.1	437	1 TMS4_HUMAN	Q9nrs4 homo sapien
2	526.5	32.9	418	1 HATT_HUMAN	O60235 homo sapien
3	523	32.7	453	1 TMS3_MOUSE	Q8k1t0 mus musculus
4	517	32.3	455	1 TMS5_MOUSE	Q9er04 mus musculus
5	509.5	31.8	454	1 TMS3_HUMAN	P57727 homo sapien
6	506	31.6	492	1 TMS2_HUMAN	O15393 homo sapien
7	505.5	31.6	417	1 HEP5_HUMAN	P05981 homo sapien
8	499	31.2	436	1 ACRO_MOUSE	P23578 mus musculus
9	492	30.8	638	1 KAL_MOUSE	P26262 mus musculus
10	491.5	30.7	436	1 HEP5_MOUSE	O35453 mus musculus
11	491	30.7	437	1 ACRO_RAT	P29293 rattus norv
12	489	30.6	638	1 KAL_RAT	P14272 rattus norv
13	487	30.4	457	1 TMS5_HUMAN	Q9h3s3 homo sapien
14	483	30.2	324	1 TEST_MOUSE	Q9jhj7 mus musculus
15	480.5	30.0	321	1 TRYG_HUMAN	Q9nrr2 homo sapien
16	479.5	30.0	416	1 HEP5_RAT	Q05511 rattus norv
17	479	29.9	638	1 KAL_HUMAN	P03952 homo sapien
18	476	29.8	421	1 ACRO_HUMAN	P10323 homo sapien
19	476	29.8	421	1 TMS2_MOUSE	Q9jiq8 mus musculus
20	475.5	29.7	490	1 ACRO_MOUSE	Q9ql7 mus musculus
21	473.5	29.6	415	1 ACRO_PIG	P08001 sus scrofa
22	470.5	29.4	246	1 TRY2_MOUSE	P07146 mus musculus
23	469	29.3	277	1 TRY2_ANOGA	P35036 anopheles g
24	464.5	29.0	247	1 TRY2_HUMAN	P07478 homo sapien
25	463.5	29.0	246	1 TRY1_RAT	P00762 rattus norv
26	463.5	29.0	267	1 TRY7_ANOGA	P35041 anopheles g
27	463.5	29.0	431	1 ACRO_RABIT	P48038 oryctolagus
28	461.5	28.8	422	1 DES1_HUMAN	Q9ul52 homo sapien
29	459.5	28.7	247	1 TRY2_CANFA	P06872 canis famul
30	458	28.6	1069	1 ENTK_MOUSE	P97435 mus musculus
31	457	28.6	342	1 PSS8_RAT	Q9es87 rattus norv
32	456.5	28.5	275	1 TRY3_ANOGA	P35037 anopheles g
33	451.5	28.2	246	1 TRY2_RAT	P00763 rattus norv

RESULT 1					
TMS4_HUMAN					
ID	TMS4_HUMAN	STANDARD;	PRT;	437	AA.
AC	Q9NRS4; Q9NZA5;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	15-SEP-2003 (Rel. 42, Last annotation update)				
DE	Transmembrane protease, serine 4 (EC 3.4.21.-) (Membrane-type serine				
DE	protease 2) (MT-SP2).				
GN	TMPSR54 OR TMPSR53.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	{1}				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Pancratic carcinoma;				
RX	MEDLINE=20283276; PubMed=10825129;				
RA	Wallrapp C., Haehnel S., Mueller-Pillasch F., Burghardt B.,				
RA	Iwamura T., Ruthenbuenger M., Lerch M.M., Adler G., Gress T.M.;				
RT	"A novel transmembrane serine protease (TMPSR53) overexpressed in				
RT	pancreatic cancer.";				
RL	Cancer Res. 60:2602-2606(2000).				
RN	{2}				
RP	SEQUENCE FROM N.A.				
RA	Smeekens S.S., Lorimer D.D., Wang E., Hou J., Linnevers C.;				
RT	"MT-SP2, a novel type II membrane serine protease expressed in				
RT	trachea, colon, and small intestine: identification, cloning, and				
RT	chromosomal localization.";				
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.				
RN	{3}				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Pancreas;				
RX	MEDLINE=22388257; PubMed=12477932;				
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,				
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,				
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,				
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,				
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,				
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,				
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,				
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,				
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,				
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,				
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA	Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,				
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,				
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,				
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,				
RA	Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;				
RT	"Generation and initial analysis of more than 15,000 full-length				
RT	human and mouse cDNA sequences.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).				
CC	-!- FUNCTION: MAY PLAY A ROLE IN TUMOR PROGRESSION, METASTASIS				
CC	FORMATION AND TUMOR INVASION.				

Q9y6m0 homo sapien
P07477 homo sapien
Q9esd1 mus musculu
P20918 mus musculu
P56677 mus musculu
P35035 anopheles g
P35030 homo sapien
Q9Y5Y6 homo sapien
P00760 bos taurus
Q29463 bos taurus
P06867 sus scrofa
P00761 sus scrofa

ALIGNMENTS

C -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
C -!- TISSUE SPECIFICITY: HIGH LEVELS IN PANCREATIC, GASTRIC, COLORECTAL
C AND AMPULLARY CANCER. VERY WEAK EXPRESSION IN NORMAL
C GASTROINTESTINAL AND UROGENITAL TRACT.
C -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
C -!- SIMILARITY: Contains 1 SRCR domain.
C -!- SIMILARITY: Contains 1 LDL-receptor class A domain.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
C or send an email to license@isb-sib.ch).
C -----
R EMBL; AF179224; AAF74526.1; --
R EMBL; AF216312; AAF31436.1; --
R EMBL; BC011703; AAH11703.1; --
R HSSP; P00763; LDPO.
R MEROPS; S01.034; --
R Genew; HGNC:11878; TMPRSS4.
R MIM; 606565; --
R GO; GO:0016021; C:integral to membrane; NAS.
R GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
R InterPro; IPR001314; Chymotrypsin.
R InterPro; IPR002172; LDL_receptor_A.
R InterPro; IPR001254; Ser_protease_Try.
R InterPro; IPR001190; Sscr_receptor.
R Pfam; PF00057; ldl_recept_a; 1.
R Pfam; PF00089; trypsin; 1.
R PRINTS; PR00722; CHYMOTRYPSIN.
R SMART; SM00192; LDLa; 1.
R SMART; SM00202; SR; 1.
R SMART; SM00020; Tryp_SPC; 1.
R PROSITE; PS01209; LDLRA_1; FALSE_NEG.
R PROSITE; PS50068; LDLRA_2; FALSE_NEG.
R PROSITE; PS00420; SRCR_1; FALSE_NEG.
R PROSITE; PS00287; SRCR_2; 1.
R PROSITE; PS50240; TRYPSIN_DOM; 1.
R PROSITE; PS00134; TRYPSIN_HIS; 1.
R PROSITE; PS00135; TRYPSIN_SER; 1.
W Hydrolase; Serine protease; Transmembrane; Signal-anchor.
T DOMAIN 1 38 CYTOPLASMIC (POTENTIAL).
T TRANSMEM 39 59 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
T DOMAIN 60 437 EXTRACELLULAR (POTENTIAL).
T DOMAIN 61 93 LDL-RECEPTOR CLASS A.
T DOMAIN 94 194 SRCR.
T DOMAIN 205 437 SERINE PROTEASE.
T ACT_SITE 245 245 CHARGE RELAY SYSTEM (BY SIMILARITY).
T ACT_SITE 290 290 CHARGE RELAY SYSTEM (BY SIMILARITY).
T ACT_SITE 387 387 CHARGE RELAY SYSTEM (BY SIMILARITY).
T SITE 204 205 CLEAVAGE (POTENTIAL).
T DISULFID 64 83 BY SIMILARITY.
T DISULFID 77 92 BY SIMILARITY.
T DISULFID 127 183 BY SIMILARITY.
T DISULFID 140 193 BY SIMILARITY.
T DISULFID 196 310 BY SIMILARITY.
T DISULFID 230 246 BY SIMILARITY.
T DISULFID 356 372 BY SIMILARITY.
T DISULFID 383 410 BY SIMILARITY.
T CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CONFLICT 1 31 MLQPPDSQPLNSLDVKPLRKPRIPMETFRK -> MSNPCA
NPVSPWRPSES (IN REF. 2).
Q SEQUENCE 437 AA; 48204 MW; 351B2FD4A8657B12 CRC64;
Query Match 79.1%; Score 1266; DB 1; Length 437;
Best Local Similarity 97.4%; Pred. No. 5.5e-111;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Y 51 KIVGGYALDVDSWPMQVSIQYDKQHVCGSILDPHWVLTAAHCFRKHTDVFNWVKVRAGSD 110

Db 204 RVVGGEASVDSWPMQVSIQYDKQHVCGSILDPHWVLTAAHCFRKHTDVFNWVKVRAGSD 263
QY 111 KLGSPSLAVAKIIIIIEFNMPKNDIALMKLPFLTFSGTVRPICLPFFDELTPTATP 170
Db 264 KLGSPSLAVAKIIIIIEFNMPKNDIALMKLPFLTFSGTVRPICLPFFDELTPTATP 323
QY 171 LWIIGWFTKQGGKMSDILLQASVQVIDSTRNADDAAYQGEVTERKMCAGIPEGGVDTG 230
Db 324 LWIIGWFTKQGGKMSDILLQASVQVIDSTRNADDAAYQGEVTERKMCAGIPEGGVDTG 383
QY 231 QGDSGGPLMYQSDQHVHVGVISWGYGCGGPSTPGVYTKVSAYLNWIYNVWKAEL 284
Db 384 QGDSGGPLMYQSDQHVHVGVISWGYGCGGPSTPGVYTKVSAYLNWIYNVWKAEL 437
RESULT 2
HATT_HUMAN
ID HATT_HUMAN STANDARD; PRT; 418 AA.
AC O60235;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Airway trypsin-like protease precursor (EC 3.4.21.-).
GN HAT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98234382; PubMed=9565616;
RA Yamaoka K., Masuda K.-I., Ogawa H., Takagi K., Umemoto N., Yasuoka S.;
RT "Cloning and characterization of the cDNA for human airway trypsin-
like protease.";
RL J. Biol. Chem. 273:11895-11901(1998).
RN [2]
RP SEQUENCE OF 187-206, AND CHARACTERIZATION.
RX MEDLINE=97224034; PubMed=9070615;
RA Yasuoka S., Ohnishi T., Kawano S., Tsuchihashi S., Ogawara M.,
RA Masuda K.-I., Yamaoka K., Takahashi M., Sano T.;
RT "Purification, characterization, and localization of a novel
trypsin-like protease found in the human airway.";
RL Am. J. Respir. Cell Mol. Biol. 16:300-308(1997).
CC -!- FUNCTION: May play some biological role in the host defense system
on the mucous membrane independently of or in cooperation with
other substances in airway mucous or bronchial secretions.
CC -!- CATALYTIC ACTIVITY: Preferentially cleaves the C-terminal side of
arginine residues at the P1 position of certain peptides, cleaving
Boc-Phe-Ser-Arg-4-methylcoumaryl-7-amide most efficiently and
having an optimum pH of 8.6 with this substrate.
CC -!- ENZYME REGULATION: Strongly inhibited by diisopropyl
fluorophosphate, leupeptin, antipain, aprotinin, and soybean
trypsin inhibitor, but hardly inhibited by secretory leukocyte
protease inhibitor at 10 microm.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY
CLEAVAGE AND SECRETED.
CC -!- TISSUE SPECIFICITY: Located in the cells of the submucosal serous
glands of the bronchi and trachea.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB002134; BAA28691.1; --
DR HSSP; P00750; 1RTF.

```

R MEROPS; S01.301; -.
R MIM; 605369; -.
R GO; GO:0005576; C:extracellular; TAS.
R GO; GO:0005887; C:integral to plasma membrane; TAS.
R GO; GO:0008233; F:peptidase activity; TAS.
R GO; GO:0007585; P:respiratory gaseous exchange; TAS.
R InterPro; IPR001314; Chymotrypsin.
R InterPro; IPR000082; SEA domain.
R InterPro; IPR001254; Ser_protease_Try.
R Pfam; PF01390; SEA; 1.
R Pfam; PF00089; trypsin; 1.
R PRINTS; PR00722; CHYMOTRYPSIN.
R SMART; SM00200; SEA; 1.
R SMART; SM00020; Tryp_Spc; 1.
R PROSITE; PS50024; SEA; 1.
R PROSITE; PS50240; TRYPSIN_DOM; 1.
R PROSITE; PS00134; TRYPSIN_HIS; 1.
R PROSITE; PS00135; TRYPSIN_SER; 1.
W Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
W Glycoprotein. 1 186 AIRWAY TRYPSIN-LIKE PROTEASE, NON-
CHAIN CATALYTIC CHAIN.
T T CHAIN 187 418 AIRWAY TRYPSIN-LIKE PROTEASE, CATALYTIC
CHAIN.
T T DOMAIN 1 20 CYTOPLASMIC (POTENTIAL).
T T TRANSMEM 21 41 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
T T DOMAIN 42 418 EXTRACELLULAR (POTENTIAL).
T T DOMAIN 44 164 SEA.
T T DOMAIN 187 417 SERINE PROTEASE.
T T ACT_SITE 227 227 CHARGE RELAY SYSTEM (BY SIMILARITY).
T T ACT_SITE 272 272 CHARGE RELAY SYSTEM (BY SIMILARITY).
T T ACT_SITE 368 368 CHARGE RELAY SYSTEM (BY SIMILARITY).
T T DISULFID 173 292 INTERCHAIN (BY SIMILARITY).
T T DISULFID 212 228 BY SIMILARITY.
T T DISULFID 337 353 BY SIMILARITY.
T T DISULFID 364 393 BY SIMILARITY.
T T CARBOHYD 144 144 N-LINKED (GLCNAC...) (POTENTIAL).
Q SEQUENCE 418 AA; 46263 MW; F4BCLDB020CFBBD0 CRC64;

```

Query Match 32.9%; Score 526.5; DB 1; Length 418;
Best local Similarity 41.8%; Pred. No. 9.6e-42;
Matches 97; Conservative 49; Mismatches 79; Indels 7; Gaps 4;

```

Y 49 DDKIVGGYALDVDSWPQVSIQYDKQHVCGGSLDPHWLTAACHFRKHTDVFNWKVRAG 108
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
b 184 EQRILGGTEAEGSWPQVSLRLNNAHCGGSLINNMWILTAACHFRSNSNPRDWIATSG 243
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 109 SDKLGSPSL--AVAKIIIEFNEMYPKNDIALMKLPPLTFSGTVRPICLPFFDELT 166
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
b 244 IST--TFPKLRMRVRNLIHNHYKSATHENDIALVRLNSVTFTKDIHSVCLPAATQNP 301
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 167 PATPLWIIIGFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTERKMCAGIPEGG 226
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
b 302 PGSTAYVTGWGAQEYAGHTVPE-LRQGVRIISNDVCNAPHSYNGAILSGMLCAGVPQGG 360
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 227 VDTCCQDGGGLMYQSDQ--WHVVGVISWGYCGGSPSPGVYTKVSAYLNWI 276
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
b 361 VDACCQDGGGLVQEDSRRLWFIVGVISWGDQCGLPDKPGVYTRVTAYLDWI 412
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

ESULT 3
MS3_MOUSE
D_TMS3_MOUSE STANDARD; PRT; 453 AA.
C Q8KJTO; Q8VDE0;
T 15-SEP-2003 (Rel. 42, Created)
T 15-SEP-2003 (Rel. 42, Last sequence update)
T 15-SEP-2003 (Rel. 42, Last annotation update)
E Transmembrane protease, serine 3 (EC 3.4.21.-).
N TMPSRS3.
S Mus musculus (Mouse).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND FUNCTION IN ENAC
RP CLEAVAGE.
RX PubMed=12393794;
RA Guipponi M., Vuagniaux G., Wattenhofer M., Shibuya K., Vazquez M.,
RA Dougherty L., Scamuffa N., Guida E., Okui M., Rossier C., Hancock M.,
RA Buchet K., Raymond A., Hummler E., Marzella P.L., Kudoh J.,
RA Shimizu N., Scott H.S., Antonarakis S.E., Rossier B.C.;
RT "The transmembrane serine protease (TMPSRS3) mutated in deafness
RL Hum Mol Genet. 11:2829-2836(2002).
CC -!- FUNCTION: Probable protease. Seems to be capable of activating
CC ENAC.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
CC reticulum.
CC -!- TISSUE SPECIFICITY: Expressed in the spiral ganglion, the cells
CC supporting the organ of Corti and the stria vascularis.
CC -!- PTM: Undergoes autolytic activation.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: Contains 1 SRCR domain.
CC -!- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ429216; CAD22137.1; -.
DR EMBL; AJ300738; CAC83350.1; -.
DR HSSP; P00761; 1AN1.
DR MGD; MGI:2155445; Tmprss3.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS50068; LDLRA_2; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
KW Endoplasmic reticulum.
FT DOMAIN 1 48
FT TRANSMEM 49 69
FT -----
FT DOMAIN 70 453
FT DOMAIN 72 108 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 104 205 LDL-RECEPTOR CLASS A.
FT DOMAIN 217 448 SRCR.
FT ACT_SITE 257 257 SERINE PROTEASE.
FT ACT_SITE 304 304 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 400 400 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 216 217 CLEAVAGE (POTENTIAL).
FT DISULFID 73 85 BY SIMILARITY.
FT DISULFID 79 98 BY SIMILARITY.
FT DISULFID 92 107 BY SIMILARITY.
FT DISULFID 129 194 BY SIMILARITY.
FT DISULFID 142 204 BY SIMILARITY.
FT DISULFID 207 324 BY SIMILARITY.
FT DISULFID 242 258 BY SIMILARITY.
FT DISULFID 338 406 BY SIMILARITY.
FT DISULFID 369 385 BY SIMILARITY.

```

T DISULFID 396 424 BY SIMILARITY.
T CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CONFLICT 117 117 L -> H (IN REF. 1; CAC83350).
T CONFLICT 246 246 V -> I (IN REF. 1; CAC83350).
Q SEQUENCE 453 AA; 49491 MW; 1ABCBF10AF61EF6 CRC64;
Query Match 32.7%; Score 523; DB 1; Length 453;
Best Local Similarity 44.2%; Pred. No. 2.3e-41;
Matches 106; Conservative 41; Mismatches 83; Indels 10; Gaps 6;
Y 51 KIVGGYALDVDSWPQVSIQYDKQHVCGSILDPHWLTAHCFRKHDTDFNWKVRAGSD 110
b 216 RIVGGNMSSLTQWPQVSLQFGYHLCGGSVITPLWIVTAHCVYDLYHPKSWTVQVGLV 275
Y 111 KL--GSFPSLAVAKIIIEFNPMY-PK--DNIDIALMKLOFPLTFSGTVRPICLPFFDEEL 165
b 276 SLMDSPVPVSHLVEKII---YHSKYKPKRLGNDIALMKLSEPLTFDETIQPICLPNSENF 332
Y 166 TPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAVQGEVTEKMMCAPIEG 225
b 333 PDGKLCWTSGWGAT-EDGDASPVNLNHAAPLISNKKICNHRDVYGGIISPSMLCAGYLKG 391
Y 226 GVDTCQDGGPLMVQSDQ-WHVVGIVSWGCGGPGSTPGVYTKVSAYLNIWYVWKAEL 284
b 392 GVDSCQDGGPLVCQERRLWKLVGATSGFGICAEVKNKPGVYTRITSFLDWIHEQLERDL 451
RESULT 4
MS5 MOUSE STANDARD; PRT; 455 AA.
D TMS5 MOUSE Q9ER02; Q9ER03;
C Q9ER04; Q9ER02; Q9ER03;
T 16-OCT-2001 (Rel. 40, Created)
T 16-OCT-2001 (Rel. 40, Last sequence update)
T 28-FEB-2003 (Rel. 41, Last annotation update)
E Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
N TMPSR55.
S Mus musculus (Mouse).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
X NCBI_TaxID=10090;
N [1]
P SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
C TISSUE=Brain;
A Mitsui S., Yamaguchi N.;
T "Molecular cloning of mouse spinesin."
L Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
C -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
C -!- ALTERNATIVE PRODUCTS:
C Event=Alternative splicing; Named isoforms=4;
C Name=4;
C IsoId=Q9ER04-1; Sequence=Displayed;
C Name=1;
C IsoId=Q9ER04-2; Sequence=VSP_005397, VSP_005398;
C Name=2;
C IsoId=Q9ER04-3; Sequence=VSP_005395;
C Name=3;
C IsoId=Q9ER04-4; Sequence=VSP_005396;
C -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
C -!- SIMILARITY: Contains 1 SRCR domain.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
C or send an email to license@isb-sib.ch).

CC EMBL; AB016229; BAB20276.1; --
DR EMBL; AB016230; BAB20277.1; --
DR EMBL; AB016423; BAB20278.1; --
DR EMBL; AB041037; BAB40328.1; --
DR HSSP; P00763; IDPO.
DR MEROPS; S01.313; --
DR MGD; MGI:1933407; TmpRSS5.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS0287; SRCR_2; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor;
KW Glycoprotein; Alternative splicing.
FT DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 71 455 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 112 207 SRCR.
FT DOMAIN 218 455 SERINE PROTEASE.
FT ACT_SITE 258 258 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 217 218 CLEAVAGE (POTENTIAL).
FT DISULFID 135 196 BY SIMILARITY.
FT DISULFID 148 206 BY SIMILARITY.
FT DISULFID 209 328 BY SIMILARITY.
FT DISULFID 243 259 BY SIMILARITY.
FT DISULFID 374 390 BY SIMILARITY.
FT DISULFID 401 429 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 144 Missing (in isoform 2).
FT VARSPLIC 1 10 /FTId=VSP_005395.
FT VARSPLIC 1 182 /FTId=VSP_005396.
FT VARSPLIC 183 192 /FTId=VSP_005397.
FT CONFLICT 325 325 GGLVEEAWKP -> MEAQVGLLWV (in isoform 1).
FT SEQUENCE 455 AA; 49632 MW; 5CFC31789C6899AA CRC64;
Query Match 32.3%; Score 517; DB 1; Length 455;
Best Local Similarity 42.4%; Pred. No. 8.3e-41;
Matches 101; Conservative 41; Mismatches 82; Indels 14; Gaps 6;
QY 51 KIVGGYALDVDSWPQVSIQYDKQHVCGSILDPHWLTAHCFRKHDTDFNWKVR. 107
Db 217 RIVGGQAVASGRWPQVSIQYDKQHVCGSILDPHWLTAHCFRKHDTDFNWKVR. 275
QY 108 GSDKLGS---FPSLAVAKIIIEFNPMYKDN---DIALMKLOFPLTFSGTVRPICLPFF 161
Db 276 GLVSHGAVRQHOGTWEKIIIP---HPLYSAQNHVDYVALLQLRTPNPSDTVDVCLPAK 332
QY 162 DEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAVQGEVTEKMMCA 221
Db 333 EQYFPWGSQCWVSGWGHTDPSHTSSDTLQDMVPLLSHLNCSNCMYSGALTHRLMCA 392
QY 222 IPEGGVDTCCQDGGPLMVQSDQ-WHVVGIVSWGCGGPGSTPGVYTKVSAYLNIWY 278
Db 393 YLDGRADACQDGGPLMVQSDQ-WHVVGIVSWGCGGPGSTPGVYTKVSAYLNIWY 450
RESULT 5
TMS3_HUMAN

D TMS3 HUMAN STANDARD; PRT; 454 AA.
C P57727;
T 16-OCT-2001 (Rel. 40, Created)
T 16-OCT-2001 (Rel. 40, Last sequence update)
T 15-SEP-2003 (Rel. 42, Last annotation update)
E Transmembrane protease, serine 3 (EC 3.4.21.-) (Serine protease
E TADG-12) (Tumor associated differentially-expressed gene-12 protein).
N TMPRSS3 OR TADG12 OR ECHOS1.
S Homo sapiens (Human).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
X NCBI_TaxID=9606;
N [1]
P SEQUENCE FROM N.A. (ISOFORMS A AND T).
C TISSUE=Ovarian carcinoma;
X MEDLINE=20521358; PubMed=11068177;
A Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N.,
A Wang Y., Parmley T.H., O'Brien T.J.,
T "Ovarian tumor cells express a novel multi-domain cell surface serine
T protease."
L Biochim. Biophys. Acta 1502:337-350(2000).
N [2]
P SEQUENCE FROM N.A. (ISOFORMS A; B AND D), AND VARIANT ILE-53.
X MEDLINE=20578749; PubMed=11137999;
A Scott H.S., Kudoh J., Wattenhofer M., Shibuya K., Berry A., Chrast R.,
A Guipponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,
A Younus F., Mehdi S.Q., Radhakrishna U., Papasavvas M.P., Gehrig C.,
A Rossier C., Korostishevsky M., Gal A., Shimizu N., Bonne-Tamir B.,
A Antonarakis S.E.;
T "Insertion of beta-satellite repeats identifies a transmembrane
T protease causing both congenital and childhood onset autosomal
T recessive deafness."
L Nat. Genet. 27:59-63(2001).
N [3]
P SUBCELLULAR LOCATION, AND FUNCTION IN ENAC CLEAVAGE.
X PubMed=12393794;
A Guipponi M., Vuagniaux G., Wattenhofer M., Shibuya K., Vazquez M.,
A Dougherty L., Scamuffa N., Guida E., Okui M., Rossier C., Hancock M.,
A Buchet K., Raymond A., Hummler E., Marzella P.L., Kudoh J.,
A Shimizu N., Scott H.S., Antonarakis S.E., Rossier B.C.;
T "The transmembrane serine protease (TMPRSS3) mutated in deafness
T DFNB8/10 activates the epithelial sodium channel (ENAC) in vitro."
L Hum. Mol. Genet. 11:2829-2836(2002).
N [4]
P VARIANTS DFNB8/DFNB10 CYS-251 AND LEU-404.
X MEDLINE=21354482; PubMed=11462234;
A Masmoudi S., Antonarakis S.E., Schwede T., Ghorbel A.M., Gratri M.,
A Pappasavas M.P., Drira M., Elgaied-Boulila A., Wattenhofer M.,
A Rossier C., Scott H.S., Ayadi H., Guipponi M.;
T "Novel missense mutations of TMPRSS3 in two consanguineous Tunisian
T families with non-syndromic autosomal recessive deafness."
L Hum. Mutat. 18:101-108(2001).
N [5]
P VARIANTS DFNB8/DFNB10 TRP-109; PHE-194 AND ARG-407, AND VARIANTS
P ILE-53; SER-111 AND VAL-253.
X MEDLINE=21317610; PubMed=11424922;
A Ben-Yosef T., Wattenhofer M., Riazuddin S., Ahmed Z.M., Scott H.S.,
A Kudoh J., Shibuya K., Antonarakis S.E., Bonne-Tamir B.,
A Radhakrishna U., Naz S., Ahmed Z., Riazuddin S., Pandya A.,
A Nance W.E., Wilcox E.R., Friedman T.B., Morell R.J.;
T "Novel mutations of TMPRSS3 in four DFNB8/B10 families segregating
T congenital autosomal recessive deafness."
L J. Med. Genet. 38:396-400(2001).
N [6]
P VARIANT DFNB8/DFNB10 GLY-103, AND VARIANTS ASN-173 AND THR-426.
X MEDLINE=21904597; PubMed=11907649;
A Wattenhofer M., Di Iorio V., Rabionet R., Dougherty L., Pampanos A.,
A Schwede T., Montserrat-Sentis B., Arbones L., Iliades T.,
A Pasquadibisceglie A., D'Amelio M., Alwan S., Rossier C., Dahl H.-H.M.,
A Petersen M.B., Estivill X., Gasparini P., Scott H.S.,
A Antonarakis S.E.;
T "Mutations in the TMPRSS3 gene are a rare cause of childhood
T nonsyndromic deafness in Caucasian patients."
J. Mol. Med. 80:124-131(2002).
-!- FUNCTION: Probable protease. Seems to be capable of activating
ENAC.
-!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
reticulum.
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=4;
Name=A;
IsoId=P57727-1; Sequence=Displayed;
Name=B; Synonyms=C;
IsoId=P57727-2; Sequence=VSP_005391;
Name=D;
IsoId=P57727-3; Sequence=VSP_005392;
Name=T; Synonyms=Truncated, TADG-12V;
IsoId=P57727-4; Sequence=VSP_005393, VSP_005394;
-!- TISSUE SPECIFICITY: Expressed in many tissues including fetal
cochlea. Isoform T is found at increased levels in some
carcinomas.
-!- PTM: Undergoes autolytic activation.
-!- DISEASE: Defects in TMPRSS3 are a cause of childhood-onset
autosomal recessive neurosensory deafness 8 (DFNB8) [MIM:601072].
-!- DISEASE: Defects in TMPRSS3 are a cause of congenital autosomal
recessive neurosensory deafness 10 (DFNB10) [MIM:605316].
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
-!- SIMILARITY: Contains 1 SRCR domain.
-!- SIMILARITY: Contains 1 LDL-receptor class A domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).

EMBL; AF201380; AAC37012.1; -
EMBL; AB038157; BAB20077.1; -
EMBL; AB038158; BAB20078.1; -
EMBL; AB038159; BAB20079.1; -
EMBL; AB038160; BAB20080.1; -
HSSP; P00763; LDPO.
MEROPS; S01.079; -.
Genew; HGNC:11877; TMPRSS3.
MIM; 605511; -
MIM; 601072; -
MIM; 605316; -
GO; GO:0016021; C:integral to membrane; NAS.
GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
InterPro; IPR001314; Chymotrypsin.
InterPro; IPR002172; LDL_receptor_A.
InterPro; IPR001254; Ser_protease_Try.
InterPro; IPR001190; Srcr_receptor.
Pfam; PF00057; ldl_recept_a; 1.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SM00192; LDLa; 1.
SMART; SM00202; SR; 1.
SMART; SM00020; Tryp_Spc; 1.
PROSITE; PS01209; LDLRA_1; 1.
PROSITE; PS00068; LDLRA_2; 1.
PROSITE; PS00420; SRCR_1; FALSE_NEG.
PROSITE; PS00287; SRCR_2; 1.
PROSITE; PS0240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SER; 1.
Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
Endoplasmic reticulum; Deafness; Alternative splicing;
Disease mutation; Polymorphism.
DOMAIN 1 48 CYTOPLASMIC (POTENTIAL).
TRANSMEM 49 69 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
DOMAIN 70 454 EXTRACELLULAR (POTENTIAL).
DOMAIN 72 108 LDL-RECEPTOR CLASS A.


```
R InterPro; IPR001314; Chymotrypsin.
R InterPro; IPR002172; LDL_receptor_A.
R InterPro; IPR001254; Ser_protease_Try.
R InterPro; IPR001190; Srcr_receptor.
R Pfam; PF00089; trypsin; 1.
R PRINTS; PR00722; CHYMOTRYPSIN.
R SMART; SM00192; LDLa; 1.
R SMART; SM00202; SR; 1.
R SMART; SM00020; Tryp_SPC; 1.
R PROSITE; PS01209; LDLRA_1; 1.
R PROSITE; PS50068; LDLRA_2; 1.
R PROSITE; PS00420; SRCR_1; FALSE_NEG.
R PROSITE; PS50287; SRCR_2; 1.
R PROSITE; PS50240; TRYPSIN_DOM; 1.
R PROSITE; PS50240; TRYPSIN_HIS; 1.
R PROSITE; PS00134; TRYPSIN_SER; 1.
R PROSITE; PS00135; TRYPSIN_SER; 1.
W Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
W Polymorphism.
T CHAIN 1 255 TRANSMEMBRANE PROTEASE, SERINE 2, NON-
T CHAIN 256 492 CATALYTIC CHAIN.
T CHAIN 84 105 TRANSMEMBRANE PROTEASE, SERINE 2,
T CHAIN 84 105 CATALYTIC CHAIN.
T DOMAIN 1 84 CYTOPLASMIC (POTENTIAL).
T TRANSMEM 85 105 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
T (POTENTIAL).
T DOMAIN 106 492 EXTRACELLULAR (POTENTIAL).
T DOMAIN 112 149 LDL-RECEPTOR CLASS A.
T DOMAIN 150 242 SRCR.
T DOMAIN 256 492 SERINE PROTEASE.
T ACT_SITE 296 296 CHARGE RELAY SYSTEM (BY SIMILARITY).
T ACT_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
T ACT_SITE 441 441 CHARGE RELAY SYSTEM.
T SITE 255 256 CLEAVAGE (POTENTIAL).
T DISULFID 113 126 BY SIMILARITY.
T DISULFID 120 139 BY SIMILARITY.
T DISULFID 133 148 BY SIMILARITY.
T DISULFID 172 231 BY SIMILARITY.
T DISULFID 185 241 BY SIMILARITY.
T DISULFID 244 365 INTERCHAIN (BY SIMILARITY).
T DISULFID 281 297 BY SIMILARITY.
T DISULFID 410 426 BY SIMILARITY.
T DISULFID 437 465 BY SIMILARITY.
T CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
T VARIANT 449 449 K -> N (IN dBSNP:1056602).
T /FTid=VAR_011692.
T MUTAGEN 255 255 R->Q: LOSS OF CLEAVAGE.
T MUTAGEN 441 441 S->A: LOSS OF ACTIVITY.
T CONFLICT 160 160 M -> V (IN REF. 3).
T CONFLICT 242 242 I -> L (IN REF. 1).
T CONFLICT 329 329 E -> Q (IN REF. 1).
T CONFLICT 489 491 RAD -> KAN (IN REF. 1).
T SEQUENCE 492 AA; 53891 MW; CAB44FD174A9076B CRC64;
```

```
Query Match 31.6%; Score 506; DB 1; Length 492;
Best Local Similarity 43.8%; Pred. No. 9.7e-40;
Matches 106; Conservative 37; Mismatches 85; Indels 14; Gaps 7;

Y 51 KIVGGYALDVSWPQVSIQDKQHVCGGSILDPHWVLTAAHCFRKH-TDVFNWVKVRAGS 109
:||||: :||||: :||||: :||||: :||||: :||||: :||||: :||||:
b 255 RIVGGSALPGAWPQVSLHVQNVHVCGGSIIPTPEWIVTAAHCHVERPLNPNWHTAFAGI 314

Y 110 DKLGSF----PSLAVAKIIIEFNPMY---PKNDIALMKLQFPLTFSGTVRPICLPFFD 162
: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
b 315 LR-QSFMFYGAGYQVEKVI---SHPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPNPG 370

Y 163 EELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAVQGEVTEKMMCAGI 222
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 371 MMLQPEQLCWISGWGATEEK-GKTSEVLNAKVLLETQRCNSRYVDNLITPAMICAGF 429

Y 223 PEGGYDTCCGDSGGPLMYQSDQ-WHVVGIVSWGYGCGSPSTPGVYTKVSAYLNWIYNVWK 281
: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
b 430 LQGNVDSQCGDSGGPLVTSKNNIWLIGDTSWGSGCCAKAYRPGVGNVMVFTDWIYRQMR 489
```

```
OY 282 AE 283
Db 490 AD 491

RESULT 7
HEPS_HUMAN
ID HEPS_HUMAN STANDARD; PRT; 417 AA.
AC P05981;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Serine protease hepsin (EC 3.4.21.-) (Transmembrane protease, serine
GN HPN OR TMRSS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC CATALYTIC CHAIN.
RX MEDLINE=88209431; PubMed=2835076;
RA Ieytus S.P., Ioeb K.R., Hagen F.S., Kurachi K., Davie E.W.;
RT "A novel trypsin-like serine protease (hepsin) with a putative
transmembrane domain expressed by human liver and hepatoma cells.";
RL Biochemistry 27:1067-1074(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas, and Spleen;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=91358502; PubMed=1885621;
RA Tsuji A., Torres-Rosado A., Arai T., le Beau M.M., Lemons R.S.,
RA Chou S.H., Kurachi K.;
RT "Hepsin, a cell membrane-associated protease. Characterization,
tissue distribution, and gene localization.";
RL J. Biol. Chem. 266:16948-16953(1991).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=93348237; PubMed=8346233;
RA Torres-Rosado A., O'Shea K.S., Tsuji A., Chou S.H., Kurachi K.;
RT "Hepsin, a putative cell-surface serine protease, is required for
mammalian cell growth.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:7181-7187(1993).
CC -!- FUNCTION: Plays an essential role in cell growth and maintenance
of cell morphology.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC -!- TISSUE SPECIFICITY: Present in most tissues, with the highest
level in liver.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
```

C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
C or send an email to license@isb-sib.ch).

R EMBL; M18930; AAA36013.1; -.
R EMBL; X07732; CAA30558.1; -.
R EMBL; X07002; CAA30058.1; -.
R EMBL; BC025716; AAH25716.1; -.
R PIR; S00845; S00845.
R HSSP; P00763; 1DPO.
R MEROPS; S01.224; -.
R Genew; HGNC:5155; HPN.
R MIM; 142440; -.
R GO; GO:0005887; C:integral to plasma membrane; TAS.
R GO; GO:0008151; P:cell growth and/or maintenance; TAS.
R InterPro; IPR001314; Chymotrypsin.
R InterPro; IPR001254; Ser_protease_Try.
R Pfam; PF00089; trypsin; 1.
R PRINTS; PR00722; CHYMOTRYPSIN.
R SMART; SM00020; Tryp_SPC; 1.
R PROSITE; PS00240; TRYPSIN_DOM; 1.
R PROSITE; PS00134; TRYPSIN_HIS; 1.
R PROSITE; PS00135; TRYPSIN_SER; 1.
W Hydrolase; Serine protease; Transmembrane; Signal-anchor.
T T SERINE PROTEASE HEPSIN, NON-CATALYTIC
T CHAIN 1 162
T CHAIN 163 417
T DOMAIN 1 17
T TRANSMEM 18 44
T T SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
T T (POTENTIAL).
T T EXTRACELLULAR (POTENTIAL).
T T SERINE PROTEASE.
T T CHARGE RELAY SYSTEM (BY SIMILARITY).
T T CHARGE RELAY SYSTEM (BY SIMILARITY).
T T CHARGE RELAY SYSTEM (BY SIMILARITY).
T T INTERCHAIN (BY SIMILARITY).
T T BY SIMILARITY.
T T BY SIMILARITY.
T T BY SIMILARITY.
T T N-LINKED (GLCNAC...) (POTENTIAL).
T T SEQUENCE 417 AA; 45011 MW; B2086FF661E551D7 CRC64;
Query Match 31.6%; Score 505.5; DB 1; Length 417;
Best Local Similarity 41.1%; Pred. No. 8.9e-40;
Matches 101; Conservative 36; Mismatches 94; Indels 15; Gaps 4;
Y 50 DKIVGGYALDVDSWPQVSIQYDKQHVCGSILDPHWLTAHCFRKHDTDFN-WKVRAG 108
b 161 DRIVGGRTSLGRWPQVSLRYDGAHLGSGLLSGDWLTAHCFPERNRVLSRWRFAG 220
Y 109 SDKLGSFSLAVAKIIIEFNMPV-----KNDIALMKLQFPLTFSGTVRPICLPF 160
b 221 AVAQASPHGLQLGVQAVVYHGVLFRDPNSENSNDIALVHLSPLPLTEYIQPVCLPA 280
Y 161 FDEELTPATPLMIIGWGFTKQNGKMSDILLQASVQVIDSTFRCNADDAYQGEVTEKMC 220
b 281 AGQALVDGKICTVTGWGNT-QYVGQAGVLQEARVPIISNDVNCNGADFYGNQIKPKMFCA 339
Y 221 GIPEGGVDTCCGDSGGPLMYQ-----SDQWHVVGVISWVGCGGPGSTPGVYTKVSAYLNW 275
b 340 GYPEGGIDACQDSGGPFVFCEDSISRTPRWRLCGIVSWGTCALAKQPGVYTKVSDFREW 399
Y 276 IYNVWK 281
b 400 IFQAIAK 405

RESULT 8
ACRO_MOUSE
ID ACRO_MOUSE STANDARD; PRT; 436 AA.
AC P23578;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Acrosin precursor (EC 3.4.21.10).
GN ACR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91185335; PubMed=2127931;
RA Kashiwabara S., Baba T., Takada M., Watanabe K., Yano Y., Arai Y.;
RT "Primary structure of mouse proacrosin deduced from the cDNA sequence
and its gene expression during spermatogenesis.";
RL J. Biochem. 108:785-791(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92041732; PubMed=1939002;
RA Watanabe K., Baba T., Kashiwabara S., Okamoto A., Arai Y.;
RT "Structure and organization of the mouse acrosin gene.";
RL J. Biochem. 109:828-833(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90255839; PubMed=2111255;
RA Klemm U., Maier W.-M., Tsacousidou S., Adham I., Willison K.,
RA Engel W.;
RT "Mouse preproacrosin: cDNA sequence, primary structure and
postmeiotic expression in spermatogenesis.";
RL Differentiation 42:160-166(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=92147126; PubMed=1783391;
RA Kremling H., Keime S., Wilhelm K., Adham I.M., Hameister H.,
RA Engel W.;
RT "Mouse proacrosin gene: nucleotide sequence, diploid expression, and
chromosomal localization.";
RL Genomics 11:828-834(1991).
CC -!- FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZOA.
CC IT IS A SERINE PROTEASE OF TRYPSIN-LIKE CLEAVAGE SPECIFICITY, IT
CC IS SYNTHESIZED IN A ZYMOGEN FORM, PROACROSIN AND STORED IN THE
CC ACROSOME.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of Arg-|-Xaa and Lys-|-Xaa bonds;
CC preferential cleavage Arg-|-Xaa >> Lys-|-Lys >> Lys-|-Xaa.
CC -!- SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; S66245; AAB20293.1; -.
DR EMBL; S64500; AAB20293.1; JOINED.
DR EMBL; S66243; AAB20293.1; JOINED.
DR EMBL; D00754; BAA00651.1; ALT INIT.
DR EMBL; X52466; CAA36704.1; ALT_INIT.
DR EMBL; M85170; AAA40124.1; ALT_INIT.
DR EMBL; M96430; AAA37163.1; -.
DR EMBL; M96426; AAA37163.1; JOINED.
DR EMBL; M96427; AAA37163.1; JOINED.
DR EMBL; M96428; AAA37163.1; JOINED.
DR PIR; JX0172; JX0172.
DR HSSP; P00763; 1DPO.
DR MEROPS; S01.223; -.

SMART; SM00202; SR; 1.
SMART; SM00020; Tryp SPC; 1.
PROSITE; PS0240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SER; 1.
Hydrolase; Serine protease; Transmembrane; Signal-anchor;
Alternative splicing.
CHAIN 1 181
CHAIN 182 436
DOMAIN 21 36
TRANSMEM 37 63
DOMAIN 64 436
DOMAIN 182 436
ACT_SITE 222 222
ACT_SITE 276 276
ACT_SITE 372 372
DISULFID 172 296
DISULFID 207 223
DISULFID 341 357
DISULFID 368 400
CARBOHYD 131 131
VARSPIC 25 44
CONFLICT 85 85
CONFLICT 204 204
CONFLICT 214 214
CONFLICT 228 229
CONFLICT 264 264
CONFLICT 281 281
SEQUENCE 436 AA; 46787 MW; 4A0993148C620BD0 CRC64;
Query Match 30.7%; Score 491.5; DB 1; Length 436;
Best Local Similarity 40.2%; Pred. No. 1.9e-38;
Matches 99; Conservative 39; Mismatches 93; Indels 15; Gaps 4;
50 DKIVGGYALDVDSWPWQVSIQYDKQHVCGSILDEHWVLTAAHCFRKHDTDFN-WKVRAG 108
180 DRIVGGQDSSLGRWPWQVSLRYDGTGLCGSLLSGDWVLTAAHCFPERNRVLSRWRFAG 239
109 SDKLGSFPLAVAKIIIEFNPMYP-----KNDIALMKIQPLTFSGTVRPICLPF 160
240 AVARTSPHAVQLGVQAVIYHGYLPFRDPTIDENSNDIALVHSSSLPLTEYIQVCLPA 299
161 FDEELTPATPLWTIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKAMCA 220
300 AGQALVDGKVCVTGTWGNT-QFYGQAMVLQEARVPIISNEVCNSPDPFYGNQIKPRMPCA 358
221 GIPEGGVDTCCGDSGGPLMYQ-----SDQHVHVGVISWGYCGGPGSTPGVYTKVSAYLNW 275
359 GYPEGGIDACQDSGGPFVCEDSISGTSRWRLCGIVSWGTCALARKPGVYTKVTDREW 418
276 IYNVWK 281
419 IFKAIK 424

RESULT 11
CRO-RAT STANDARD; PRT; 437 AA.
P29293;
01-DEC-1992 (Rel. 24, Created)
01-DEC-1992 (Rel. 24, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Acrosin precursor (EC 3.4.21.10).
ACR.
Rattus norvegicus (Rat).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]

SEQUENCE FROM N.A.
STRAIN=Sprague-Dawley; TISSUE=Testis;
MEDLINE=92031708; PubMed=1932123;
Klemm U., Flake A., Engel W.;
"Rat sperm acrosin: cDNA sequence, derived primary structure and
phylogenetic origin.";
Biochim. Biophys. Acta 1090:270-272(1991).
-!- FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZOA.
IT IS A SERINE PROTEASE OF TRYPSIN-LIKE CLEAVAGE SPECIFICITY, IT
IS SYNTHESIZED IN A ZYMOGEN FORM, PROACROSIN AND STORED IN THE
ACROSOME.
-!- CATALYTIC ACTIVITY: Hydrolysis of Arg-|-Xaa and Lys-|-Xaa bonds;
preferential cleavage Arg-|-Xaa >> Lys-|-Xaa.
-!- SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
DISULFIDE BONDS.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X59254; CAA41947.1; ALT_SEQ.
PIR; S18407; S18407.
HSP; P00763; LDPO.
MEROPS; S01.223; -.
InterPro; IPR001314; Chymotrypsin.
InterPro; IPR001254; Ser_protease_Try.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SM00020; Tryp SPC; 1.
PROSITE; PS0240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SER; 1.
Hydrolase; Serine protease; Glycoprotein; Zymogen; Sperm; Signal.
KW SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 437 ACROSIN.
FT CHAIN 20 42 ACROSIN LIGHT CHAIN.
FT CHAIN 43 ? ACROSIN HEAVY CHAIN.
FT PROPEP ? 437 PRO-RICH.
FT DOMAIN 43 291 SERINE PROTEASE.
FT DISULFID 25 155 INTERCHAIN (BY SIMILARITY).
FT DISULFID 29 163 INTERCHAIN (BY SIMILARITY).
FT DISULFID 74 90 BY SIMILARITY.
FT DISULFID 178 247 BY SIMILARITY.
FT DISULFID 210 226 BY SIMILARITY.
FT DISULFID 237 267 BY SIMILARITY.
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (BY SIMILARITY).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. .) (BY SIMILARITY).
FT ACT_SITE 89 89 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 241 241 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 437 AA; 48279 MW; B03BF1DD0417C0D5 CRC64;
Query Match 30.7%; Score 491; DB 1; Length 437;
Best Local Similarity 39.7%; Pred. No. 2.1e-38;
Matches 114; Conservative 38; Mismatches 107; Indels 28; Gaps 9;
QY 14 LLLVSNLLLCQGVSDYKDDDDVDAAALAPFDDDD----KIVGGYALDVDSWPWQVSI 69
Db 4 MLPTVVALVAVSVA--KDNVTCGPGC-GLRFRQNPQAGIRIVGGQTSRWAWPMWVSL 60
QY 70 QY-----DKQHVCGSILDPHWVLTAAHCFRKHDTDFNWK-----VRAGSDKLGSPF 116
Db 61 QIFTSHNSRRYHACGGSLLNSHWVLTAAHCFDNKKKYDWRKLVFGAHEIEYGRNKPVEP 120
QY 117 SLA--VAKIIIEFNPMYPKNDIALMKIQPLTFSGTVRPICLPFFDEELTPATP--LW 172
Db 121 QQERYVQKIVIEHKYNAVTEGNDIALLKVTPVTCGDFVGGCLPHF-KSGPPRIPHTCY 179

Y 221 GIPEGGVDTCCGDSGGPLMYQ-SDQWHVVGIVSWGYGCGSPSTPGVYTKVSAYLNWI 276
D 565 GYKEGGIDACKDGGSLVCKHSGRWQLVGTWEGGCARKEQPGVYTKVAEYIDWI 621

RESULT 13

VS5 HUMAN

D TMS5_HUMAN STANDARD; PRT; 457 AA.

C Q9H3S3;
I 16-OCT-2001 (Rel. 40, Created)
I 16-OCT-2001 (Rel. 40, Last sequence update)
I 28-FEB-2003 (Rel. 41, Last annotation update)
E Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
N TMPRSS5.

S Homo sapiens (Human).
C Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
K NCBI_TaxID=9606;

N [1]
P SEQUENCE FROM N.A.

C TISSUE=Brain;

X PubMed=11741986;

A Yamaguchi N., Okui A., Yamada T., Nakazato H., Mitsui S.;

I "Spinesin/TMPRSS5, a novel transmembrane serine protease, cloned from

I human spinal cord."

L J. Biol. Chem. 277:6806-6812(2002).

-!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).

-!- TISSUE SPECIFICITY: Brain-specific. Predominantly expressed in

-!- neurons, in their axons, and at the synapses of motoneurons in the

-!- spinal cord.

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

-!- SIMILARITY: Contains 1 SRCR domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

R EMBL; AB028140; BAB20375.1; -.

R HSSP; P00763; LDPO.

R Genew; HGNC:14908; TMPRSS5.

R MIM; 606751; -.

R MEROPS; S01.313; -.

R InterPro; IPR001314; Chymotrypsin.

R InterPro; IPR001254; Ser_protease_Try.

R InterPro; IPR001190; Srcr_receptor.

R Pfam; PF00089; trypsin; 1.

R PRINTS; PR00722; CHYMOTRYPSIN.

R SMART; SM00020; Tryp_SPC; 1.

R PROSITE; PS50240; TRYPSIN_DOM; 1.

R PROSITE; PS00134; TRYPSIN_HIS; 1.

R PROSITE; PS00135; TRYPSIN_SER; 1.

R PROSITE; PS00420; SRCR_1; FALSE_NEG.

R PROSITE; PS50287; SRCR_2; FALSE_NEG.

W Hydrolase; Serine protease; Transmembrane; Signal-anchor;

W Glycoprotein.

I DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).

I TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)

I (POTENTIAL).

I DOMAIN 71 457 EXTRACELLULAR (POTENTIAL).

I DOMAIN 112 207 SRCR.

I DOMAIN 218 457 SERINE PROTEASE.

I ACT_SITE 258 457 CHARGE RELAY SYSTEM (BY SIMILARITY).

I ACT_SITE 308 457 CHARGE RELAY SYSTEM (BY SIMILARITY).

I ACT_SITE 405 408 CHARGE RELAY SYSTEM (BY SIMILARITY).

I SITE 217 218 CLEAVAGE (POTENTIAL).

I DISULFID 135 196 BY SIMILARITY.

I DISULFID 148 206 BY SIMILARITY.

I DISULFID 209 328 BY SIMILARITY.

FT DISULFID 243 259 BY SIMILARITY.
FT DISULFID 374 390 BY SIMILARITY.
FT DISULFID 401 429 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 457 AA; 49574 MW; 64406AB4985A2651 CRC64;

Query Match 30.4%; Score 487; DB 1; Length 457;

Best Local Similarity 39.7%; Pred. No. 5.4e-38;

Matches 94; Conservative 44; Mismatches 87; Indels 12; Gaps 5;

OY 51 KIVGGYALDVDSWPWQVSIQYDKQHVCGGSILDPHWVLTAAHCFR--KHTDVENWKCVRAG 108

Db 217 RIVGGQSVAPGRWPWQASVALGFRHTCGSVLAPRWVVTAAHCHMSFRLARLSSWRVHAG 276

OY 109 SDKLGSFPLAVAKIIIEF---NMPYKDN---DIALMKLQFPLTPSGTVRPICLPFFD 162

Db 277 ---LVSHSAVRPHQGLVERIIPHPLYSAQNHDYDVALRLQALNFSDTVGAVCLPAKE 333

OY 163 EELTPATPLWIIGWGFKQNGCKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMMCAGI 222

Db 334 QHEPKGSRCWVSGWGHTPSHSYSSDMLQDTVVPLFSTQLCNSSCVYSGALTFRMLCAGY 393

OY 223 PEGGVDTCQDSSGGPLMY-QSDQWHVVGIVSWGYGCGSPSTPGVYTKVSAYLNWIYN 278

Db 394 LDGRADACQDSSGGPLVCPDGTWRLVGVVSWGRACAEPNHPGVYAKVAEFLDWIHD 450

RESULT 14

TEST_MOUSE

ID TEST_MOUSE STANDARD; PRT; 324 AA.

AC Q9JHJ7; Q9DA14;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Testisin precursor (EC 3.4.21.-) (Trypsin 4).

GN PRSS21.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=129/Sv;

RX MEDLINE=21153229; PubMed=11231276;

RA Scarman A.L., Hooper J.D., Boucaut K.J., Sit M.-L., Webb G.C.,

RA Normyle J.F., Antalis T.M.;

RT "Organization and chromosomal localization of the murine Testisin gene

RT encoding a serine protease temporally expressed during

RT spermatogenesis.";

RL Eur. J. Biochem. 268:1250-1258(2001).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=BALB/C; TISSUE=Testis;

RX PubMed=11259427;

RA Wong G.W., Li L., Madhusudhan M.S., Krilis S.A., Gurish M.F.,

RA Rothenberg M.E., Sali A., Stevens R.L.;

RT "Trypsin 4, a new member of the chromosome 17 family of mouse serine

RT proteases.";

RL J. Biol. Chem. 276:20648-20658(2001).

RN [3]

RP SEQUENCE OF 3-324 FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - protein search, using sw model

run on: November 23, 2003, 07:45:10 ; Search time 43.7799 Seconds
(without alignments)
1721.140 Million cell updates/sec

title: US-09-607-745-9
effect score: 1600
sequence: 1 MDSKGSQKSRLLLLLVSN.....LNWIYNVWKAELSRHHHHH 292

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 830525 seqs, 258052604 residues

total number of hits satisfying chosen parameters: 830525

inimum DB seq length: 0
aximum DB seq length: 2000000000
ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : SPTREMBL 23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1266	79.1	405	4 Q96E86	Q96E86 homo sapien
2	1048	65.5	435	11 Q8VCA5	Q8VCA5 mus musculu
3	533	33.3	537	4 Q9BYE1	Q9BYE1 homo sapien
4	528	33.0	471	11 Q8CFE0	Q8CFE0 mus musculu
5	528	33.0	581	4 Q9BYE2	Q9BYE2 homo sapien
6	526.5	32.9	279	11 Q9QZ74	Q9QZ74 rattus norv
7	524	32.8	453	11 Q8VDE0	Q8VDE0 mus musculu
8	523	32.7	453	11 Q8K1T0	Q8K1T0 mus musculu
9	522	32.6	371	11 Q8CJ16	Q8CJ16 rattus norv
10	522	32.6	445	11 Q8CJ17	Q8CJ17 rattus norv
11	522	32.6	767	13 Q9DGR2	Q9DGR2 xenopus lae
12	521.5	32.6	417	11 Q8VHJ4	Q8VHJ4 rattus norv
13	519	32.4	455	11 Q8CDR0	Q8CDR0 mus musculu
14	509.5	31.8	417	11 Q8VDV1	Q8VDV1 mus musculu
15	509.5	31.8	417	11 Q8VHK8	Q8VHK8 mus musculu
16	506	31.6	492	4 Q96T73	Q96T73 homo sapien

17	492.5	30.8	799	11 Q9DBI0	Q9dbi0 mus musculu
18	487.5	30.5	329	6 Q9GL10	Q9gl10 ovis aries
19	486	30.4	638	11 Q8ROP5	Q8rop5 mus musculu
20	483	30.2	490	11 Q920K3	Q920k3 rattus norv
21	481.5	30.1	377	6 P79343	P79343 bos taurus
22	480.5	30.0	423	11 Q8EM10	Q8em10 mus musculu
23	478.5	29.9	321	4 Q96RZ8	Q96rz8 homo sapien
24	478.5	29.9	421	11 Q60491	Q60491 cavia porce
25	477.5	29.8	416	11 Q8BZ30	Q8bz30 mus musculu
26	477.5	29.8	416	11 Q8BZ13	Q8bz13 mus musculu
27	477.5	29.8	802	4 Q8IUE2	Q8iue2 homo sapien
28	477.5	29.8	811	4 Q8IU80	Q8iu80 homo sapien
29	473.5	29.6	415	6 Q29015	Q29015 sus sp. pre
30	470	29.4	1524	13 Q91674	Q91674 xenopus lae
31	469.5	29.3	812	11 Q9ROW3	Q9row3 rattus norv
32	468	29.2	643	6 Q97506	Q97506 sus scrofa
33	467	29.2	439	11 Q8BHM9	Q8bhm9 mus musculu
34	464.5	29.0	239	4 Q8NIC9	Q8nic9 homo sapien
35	463	28.9	572	11 Q8BIK6	Q8bik6 mus musculu
36	459.5	28.7	310	11 Q9QYZ9	Q9qyz9 mus musculu
37	458	28.6	260	13 Q9W7P9	Q9w7p9 paralichthy
38	458	28.6	274	5 Q17086	Q17086 anopheles s
39	455.5	28.5	322	11 Q920S2	Q920s2 mus musculu
40	455	28.4	284	4 Q8NF86	Q8nf86 homo sapien
41	454.5	28.4	310	11 Q91XC4	Q91xc4 mus sapien
42	454	28.4	327	4 Q8N171	Q8n171 homo sapien
43	453	28.3	339	11 Q99L44	Q99l44 mus musculu
44	452.5	28.3	247	4 Q8NHM4	Q8nhm4 homo sapien
45	452.5	28.3	257	11 Q8BZ04	Q8bz04 mus musculu

ALIGNMENTS

RESULT 1
Q96E86
ID Q96E86 PRELIMINARY; PRT; 405 AA.
AC Q96E86;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to transmembrane protease, serine 4 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; BC012752; AAH12752.1; -.
DR HSSP; P00761; 1AN1.
DR MEROPS; S01.034; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL receptor A.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF000057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane.
FT NON TER 1
SQ SEQUENCE 405 AA; 44474 MW; 951ACD52D9D48E04 CRC64;

Query Match 79.1%; Score 1266; DB 4; Length 405;
Best Local Similarity 97.4%; Pred. No. 4e-118;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Y 51 KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHDTDFNWKVRAGSD 110
b 172 RVVGGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHDTDFNWKVRAGSD 231

Y 111 KLGSFPPSLAVAKIIIIIEFNPMYPKNDIALMKLQFPPLTFSGTVRPICLPFFDEELTPATP 170
b 232 KLGSFPPSLAVAKIIIIIEFNPMYPKNDIALMKLQFPPLTFSGTVRPICLPFFDEELTPATP 291

Y 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAQQGEVTEKMMKAGIPEGGVDTTC 230
b 292 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAQQGEVTEKMMKAGIPEGGVDTTC 351

Y 231 QGDSGGPLMYQSDQHVGVIVSWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 284
b 352 QGDSGGPLMYQSDQHVGVIVSWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 405

RESULT 2
8VCA5 PRELIMINARY; PRT; 435 AA.
C Q8VCA5;
T 01-MAR-2002 (TREMBlrel. 20, Created)
T 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
T 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
E Similar to transmembrane protease, serine 4 (Channel-activating
E protease 2).
N TMPSR4.
S Mus musculus (Mouse).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
X NCBI_TaxID=10090;
N [1]
P SEQUENCE FROM N.A.
C TISSUE=Breast tumor;
A Strausberg R.;
L Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
N [2]
P SEQUENCE FROM N.A.
X MEDLINE=22144321; PubMed=12149280;
A Vuagniaux G., Vallet V., Jaeger N.F., Hummler E., Rossier B.C.;
T "Synergistic Activation of ENaC by Three Membrane-bound Channel-
T activating Serine Proteases (mCAP1, mCAP2, and mCAP3) and Serum- and
T Glucocorticoid-regulated Kinase (Sgk1) in Xenopus Oocytes.";
L J. Gen. Physiol. 120:191-201(2002).
C -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
R EMBL; BC021368; AAH21368.1; -.
R EMBL; AY043240; AAK85307.1; -.
R HSP; P00761; IAN1.
R MEROPS; S01.034; -.
R MGD; MGI:2384877; TmpRSS4.
R InterPro; IPR001314; Chymotrypsin.
R InterPro; IPR002172; LDL_receptor_A.
R InterPro; IPR001254; Ser_protease_Try.
R InterPro; IPR001190; Srcr_receptor.
R Pfam; PF00057; ldl_recept_a; 1.
R PRINTS; PR00722; CHYMOTRYPSIN.
R PROSITE; PS50287; SRCR_2; 1.
R PROSITE; PS50240; TRYPSIN_DOM; 1.
R PROSITE; PS00134; TRYPSIN_HIS; 1.
R PROSITE; PS00135; TRYPSIN_SER; 1.
W Hydrolase; Protease; Serine protease; Transmembrane.
Q SEQUENCE 435 AA; 47495 MW; DC52E45A43E01369 CRC64;

Query Match 65.5%; Score 1048; DB 11; Length 435;
Best Local Similarity 78.6%; Pred. No. 2.9e-96;
Matches 184; Conservative 27; Mismatches 23; Indels 0; Gaps 0;

Y 51 KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHDTDFNWKVRAGSD 110

Db 202 RWVGVEAPVDVSWPQVSIQYDKQHVCGGSILDPHWLTAHCFRKHDTDFNWKVRAGSN 261
QY 111 KLGSFPPSLAVAKIIIIIEFNPMYPKNDIALMKLQFPPLTFSGTVRPICLPFFDEELTPATP 170
Db 262 ILGNPSLPVAKIFIAEPNPLYPKKDIALVKLQMLPTFSGSVRPICLPFFDEVLPATP 321

QY 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAQQGEVTEKMMKAGIPEGGVDTTC 230
Db 322 VWVIGWFTENGKMSDMLLQASVQVIDSTRCNADDAQQGEVTEKMMKAGIPEGGVDTTC 381

QY 231 QGDSGGPLMYQSDQHVGVIVSWGYCGGPGSTPGVYTKVSAYLNWIYNVWKAE 284
Db 382 QGDSGGPLMYHSDKWQVGVIVSWGHGCGPGSTPGVYTKVTAYLNWIYNVRKSEM 435

RESULT 3
Q9BYE1 PRELIMINARY; PRT; 537 AA.
AC Q9BYE1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Mosaic serine protease.
GN MSPS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=21167393; PubMed=11267681;
RA Kim D.R., Sharmin S., Inoue M., Kido H.;
RT "Cloning and expression of novel mosaic serine proteases with and
RT without a transmembrane domain from human lung.";
RL Biochim. Biophys. Acta 1518:204-209(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AB048797; BAB39742.1; -.
DR HSP; P00763; IDPO.
DR MEROPS; S01.087; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 537 AA; 58102 MW; A39FF4E8816DAECF CRC64;

Query Match 33.3%; Score 533; DB 4; Length 537;
Best Local Similarity 44.0%; Pred. No. 1.4e-44;
Matches 106; Conservative 38; Mismatches 87; Indels 10; Gaps 5;

QY 51 KIVGGYALDVDSWPQVSIQYDKQHVCGGSILDPHWLTAHCF--RKHTDFNWKVRAG 108
Db 295 RIVGALASDSKWPQVSLHFGTTHICGTLDAQVLTAAHCFVTRKVLGKVKYAG 354

QY 109 SDKLGSFPPSLA-VAKIIIEFNPMY--PKNDIALMKLQFPPLTFSGTVRPICLPFFDEE 164
Db 355 TSNLHQLPEASIAEII--NSNYTDEEDYDIALMRLSKPLTSAHIHPACLPFHGQT 411

QY 165 LTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAQQGEVTEKMMKAGIPE 224
Db 412 FSLNETCWTGFGKTRETDKTSFPLREVQVNLIDFKKCNLYLVDSYLTPRMTCAGDLH 471

W Hydrolase; Protease; Serine protease; Signal.
T SIGNAL 1 22 POTENTIAL.
Q SEQUENCE 279 AA; 30522 MW; 1C28069DF064546 CRC64;

Query Match 32.9%; Score 526.5; DB 11; Length 279;
Best Local Similarity 40.4%; Pred. No. 2.6e-44;
Matches 110; Conservative 49; Mismatches 96; Indels 17; Gaps 8;

Y 14 LLLVSNLLLCQGVSDYKDDDDVDAALAAPFD---DDDKIYGVYALDVDSWPPQVSI 69
b 10 LLLVLSFTL---ALTD-QDTENVLTQECGARPDLITLSEERIIGGTOAETGDWPPQVSL 65

Y 70 QYDKQHVCSSILDPHWLTAACHCFRKHDTVFNWVVRAGSDKLGSPSLAVAKIIII---126
b 66 QLNNVHCGGLISNLNVLTAACHCFRSYSPNPQQWTATFGVSTIS--PRLRVVRILAH 123

Y 127 EFNPMYPKNDIALMKLOFPLTFSGTVRPICLPFFDELTDPATPLMIIGWFTKQNGKM 186
b 124 EYNSI-TRNDIAVVQLDRPVTFTRNIHRVCLPAATQNIIPDSVAYATGWSLTYGNTV 182

Y 187 SDILLOASVQVIDSTRCNADDAAYQGEVTEKMMCAGIPEGGVDTCQDSDGGPLMYQSDQ--244
b 183 TN-LQQGEVRIVSSEVCNEPAGYGGSVLPGMLCAGVRSGAVDACQDSDGGPLVQEDTRRL 241

Y 245 WHVVGIVSWGCGGPGSTPGVYTKVSAVLNWI 276
b 242 WVVVGIVSWGVCGLPNKPGVYTRVTAARNWI 273

RESULT 7
8VDE0 PRELIMINARY; PRT; 453 AA.
D Q8VDE0
C Q8VDE0;
T 01-MAR-2002 (TrEMBLrel. 20, Created)
T 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
T 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
E TMPSR33 protein.
N TMPSR33.
S Mus musculus (Mouse).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
X NCBI_TaxID=10090;
N [1]
P SEQUENCE FROM N.A.
A Guipponi M., Scamuffa N., Scott H.S., Rossier C., Antonarakis S.E.;
T "Isolation and characterization of the mouse Tmprs3 gene."
L Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
C -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
R HSSP; PJ300738; CAC83350.1; --
R MGD; MGI:2155445; Tmprs3.
R InterPro; IPR001314; Chymotrypsin.
R InterPro; IPR002172; LDL_receptor_A.
R InterPro; IPR001254; Ser_protease_Try.
R InterPro; IPR001190; Srcr_receptor.
R Pfam; PF00057; ldl_recept_a; 1.
R Pfam; PF00089; trypsin; 1.
R PRINTS; PR00722; CHYMOTRYPSIN.
R SMART; SM00192; LDLa; 1.
R PROSITE; PS01209; TRYPSIN_DOM; 1.
R PROSITE; PS50068; LDLRA_2; 1.
R PROSITE; PS50287; SRCR_2; 1.
R PROSITE; PS50240; TRYPSIN_DOM; 1.
R PROSITE; PS00134; TRYPSIN_HIS; 1.
R PROSITE; PS00135; TRYPSIN_SER; 1.
W Hydrolase; Protease; Serine protease.
Q SEQUENCE 453 AA; 49529 MW; 21E5697DC8781BD3 CRC64;

Query Match 32.8%; Score 524; DB 11; Length 453;
Best Local Similarity 44.6%; Pred. No. 9e-44;
Matches 107; Conservative 40; Mismatches 83; Indels 10; Gaps 6;

QY 51 KIVGGYALDVDSWPPQVSIQYDKQHVCSSILDPHWLTAACHCFRKHDTVFNWVVRAGSD 110
Db 216 RIVGNNMSSLTQWPQVSLQFGYHLCGGSIIITPLWIVTAACHCVYDLYHPKSWTVQVGLV 275

QY 111 KL--GSFPSLAVAKIIIEFNPMY-PK--DNDIALMKLOFPLTFSGTVRPICLPFFDEEL 165
Db 276 SLMDSPVPVSHLVEKII---YHSKYKPKRLGNDIALMKLSEPLTFDETIQICLPNSENF 332

QY 166 TPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMMCAGIPEG 225
Db 333 PDGKLCWTSGWGAT-EDGGDASPVLNHAAPVPLISNKNICNHRDVYGGIISPSMLCAGYLKG 391

QY 226 GVDTCQDSDGGPLMYQSDQ-WHVVGIVSWGVCGLPNKPGVYTKVSAVLNWIYVWKAEL 284
Db 392 GVDSCQDSDGGPLVCQERRLMKLVGATSGFVGICAEVKNPKGVYTRITSFLDWIHEQLERDL 451

RESULT 8
Q8K1T0 PRELIMINARY; PRT; 453 AA.
ID Q8K1T0
AC Q8K1T0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type II transmembrane serine protease.
GN TMPSR33.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=citb-cj7;
RA Guipponi M., Scamuffa N., Scott H.S., Antonarakis S.E.;
RT "Isolation of the mouse Tmprs3 genomic DNA sequence."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AJ429216; CAD22137.1; --
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS50068; LDLRA_2; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane.
SQ SEQUENCE 453 AA; 49491 MW; 1ABCBF10AF6E1EF6 CRC64;

Query Match 32.7%; Score 523; DB 11; Length 453;
Best Local Similarity 44.2%; Pred. No. 1.1e-43;
Matches 106; Conservative 41; Mismatches 83; Indels 10; Gaps 6;

QY 51 KIVGGYALDVDSWPPQVSIQYDKQHVCSSILDPHWLTAACHCFRKHDTVFNWVVRAGSD 110
Db 216 RIVGNNMSSLTQWPQVSLQFGYHLCGGSIIITPLWIVTAACHCVYDLYHPKSWTVQVGLV 275

QY 111 KL--GSFPSLAVAKIIIEFNPMY-PK--DNDIALMKLOFPLTFSGTVRPICLPFFDEEL 165
Db 276 SLMDSPVPVSHLVEKII---YHSKYKPKRLGNDIALMKLSEPLTFDETIQICLPNSENF 332

QY 166 TPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMMCAGIPEG 225
Db 333 PDGKLCWTSGWGAT-EDGGDASPVLNHAAPVPLISNKNICNHRDVYGGIISPSMLCAGYLKG 391

```
Y 226 GVDTCQDSCGGLMYQSDQ-WHVVGIVSWGCGGPGSTPGVYTKVSAYLNWIYNWVKAE 284
b 392 GVDSCQDSCGGLVVCQERRLKLVGATSGFGCAEVNKPVGVTITSTFSLDWIHEQLERDL 451
RESULT 9
8CJ16
D Q8CJ16 PRELIMINARY; PRT; 371 AA.
C Q8CJ16;
T 01-MAR-2003 (TrEMBLrel. 23, Created)
T 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
T 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
E Adrenal mitochondrial protease short variant.
N AMP.
S Rattus norvegicus (Rat).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
X NCBI_TaxID=10116;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=NEDH;
A Omer S., Bicknell A.B., Lowry P.J.;
T "Identification of a rat adrenal mitochondrial protease.";
L Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
R EMBL; AF537099; AAN06758.1; -.
W Protease.
Q SEQUENCE 371 AA; 40694 MW; 89A64081D9A1FE26 CRC64;

Query Match 32.6%; Score 522; DB 11; Length 371;
Best Local Similarity 42.0%; Pred. No. 1.1e-43;
Matches 100; Conservative 42; Mismatches 82; Indels 14; Gaps 6;

Y 51 KIVGGYALDVDSWPNQVSIQYDKQHVCGGSILDPHWVLTAAHC---FRKHTDVFNNKVR 107
b 133 RIVGQAVASGRWPQASVNLGSRHTCGGSLAPYVWVTAACHCMYSFRL-SRLSSWRVHA 191
Y 108 G----SDKLGSPSLAVAKIIIEFNPMYPKDN---DIALMKLQFPLTSGTVRPICLPFF 161
b 192 GLVSHSAVRQHOGTMVEKIIP---HPLYSAGNHQHDYDVALQLRTPINFSDTSAVCLPAK 248
Y 162 DEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCAG 221
b 249 EQHFPQGSQCWVGWGHDPSTHSSDTLQDTMVPLSLDLCNSSCMYSGALTHRMCLCAG 308
Y 222 IPEGGVDTCCGDSGGPLMYQS-DQWHVVGIVSWGCGGPGSTPGVYTKVSAYLNWIYN 278
b 309 YLDGRADACQDSCGGLVVCPSGDTWHLVGVVSWGRCAPENRPGVYAKVAEFLDWIHD 366
RESULT 10
8CJ17
D Q8CJ17 PRELIMINARY; PRT; 445 AA.
C Q8CJ17;
T 01-MAR-2003 (TrEMBLrel. 23, Created)
T 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
T 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
E Adrenal mitochondrial protease long variant.
N AMP.
S Rattus norvegicus (Rat).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
X NCBI_TaxID=10116;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=NEDH;
A Omer S., Bicknell A.B., Lowry P.J.;
T "Identification of a rat adrenal mitochondrial protease.";
L Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
R EMBL; AF537098; AAN06757.1; -.
W Protease.
Q SEQUENCE 445 AA; 48440 MW; BE3F56D8372ED988 CRC64;

Query Match 32.6%; Score 522; DB 11; Length 445;
```

```
Best Local Similarity 42.0%; Pred. No. 1.4e-43;
Matches 100; Conservative 42; Mismatches 82; Indels 14; Gaps 6;

Y 51 KIVGGYALDVDSWPNQVSIQYDKQHVCGGSILDPHWVLTAAHC---FRKHTDVFNNKVR 107
b 207 RIVGQAVASGRWPQASVNLGSRHTCGGSLAPYVWVTAACHCMYSFRL-SRLSSWRVHA 265
Y 108 G----SDKLGSPSLAVAKIIIEFNPMYPKDN---DIALMKLQFPLTSGTVRPICLPFF 161
b 266 GLVSHSAVRQHOGTMVEKIIP---HPLYSAGNHQHDYDVALQLRTPINFSDTSAVCLPAK 322
Y 162 DEELTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCAG 221
b 323 EQHFPQGSQCWVGWGHDPSTHSSDTLQDTMVPLSLDLCNSSCMYSGALTHRMCLCAG 382
Y 222 IPEGGVDTCCGDSGGPLMYQS-DQWHVVGIVSWGCGGPGSTPGVYTKVSAYLNWIYN 278
b 383 YLDGRADACQDSCGGLVVCPSGDTWHLVGVVSWGRCAPENRPGVYAKVAEFLDWIHD 440
RESULT 11
Q9DGR2
ID Q9DGR2 PRELIMINARY; PRT; 767 AA.
AC Q9DGR2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Embryonic serine protease-2.
GN XESP-2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20363741; PubMed=10903452;
RA Yamada K., Takabatake T., Takeshima K.;
RT "Isolation and characterization of three novel serine protease genes
RT from Xenopus laevis.";
RL Gene 252:209-216 (2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AB038497; BAB08217.1; -.
DR HSSP; P00766; 1CHG.
DR MEROPS; S01.049; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00057; ldl_recept_a; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLa; 8.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLRA_1; 8.
DR PROSITE; PS50068; LDLRA_2; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 767 AA; 86001 MW; E0566A38796DE96E CRC64;

Query Match 32.6%; Score 522; DB 13; Length 767;
Best Local Similarity 44.2%; Pred. No. 2.9e-43;
Matches 103; Conservative 40; Mismatches 84; Indels 6; Gaps 4;

Y 51 KIVGGYALDVDSWPNQVSIQYDKQHVCGGSILDPHWVLTAAHC---FRKHTDVFNNKVR 109
b 530 RIVGTFANLGNWPQVNLQYITGVLCGSLIISPKWIVTAACHCVGYSASSAGWRVFAGT 589
Y 110 DKLGSPSLA---VAKIIIEFNPMYPKDNIALMKLQFPLTSGTVRPICLPFFDEELT 166
b 590 LTKPSYNASAYFVERIVHPGKSYTYDNDIALMKLRDEITFGYTTQPVCLPNSGMFWE 649
```

Y 167 PATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMMKAGIPEGG 226
b 650 AGTTTWISGWGSTYE-GGSVSTYLQYAAIPLIDSNVCNQSYVYNGQITSSMICAGYLSGG 708
Y 227 VDTCCQDSDGGPLM-YQSDQWHVVGIVSWGVCYGGPSTPGVYTKVSAYLNWIYN 278
b 709 VDTCCQDSDGGPLVKNRGTWMLVGDTSWGDGCGARANKPGVYGNVTTFLEWIYS 761
[1]
RESULT 12
Q8VHJ4 PRELIMINARY; PRT; 417 AA.
D Q8VHJ4;
C Q8VHJ4;
T 01-MAR-2002 (TrEMBLrel. 20, Created)
T 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
T 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
E Airway trypsin-like protease.
S Rattus norvegicus (Rat).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
X NCBI_TaxID=10116;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=Wistar; TISSUE=Trachea;
A Hansen I.A.; Fassnacht M.; Hammer F.; Schammann M.; Allolio B.;
T "Cloning and characterization of RAT, the airway trypsin-like protease
T of Rattus norvegicus."
L Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
C -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
R EMBL; AF453776; AAL50817.1; -
R HSSP; P00761; 1AN1.
R MEROPS; S01.047; -
R InterPro; IPR001314; Chymotrypsin.
R InterPro; IPR000082; SEA domain.
R InterPro; IPR001254; Ser_protease_Try.
R Pfam; PF00089; trypsin; 1.
R Pfam; PF01390; SEA; 1.
R PRINTS; PR00722; CHYMOTRYPSIN.
R SMART; SM00200; SEA; 1.
R SMART; SM00200; Tryp_Spc; 1.
R PROSITE; PS50024; SEA; 1.
R PROSITE; PS50240; TRYPSIN_DOM; 1.
R PROSITE; PS00134; TRYPSIN_HIS; 1.
R PROSITE; PS00135; TRYPSIN_SER; 1.
W Hydrolase; Protease; Serine protease.
Q SEQUENCE 417 AA; 46287 MW; DB9504158B018E21 CRC64;
Query Match 32.6%; Score 521.5; DB 11; Length 417;
Best Local Similarity 42.1%; Pred. NO. 1.4e-43;
Matches 98; Conservative 44; Mismatches 82; Indels 9; Gaps 5;
Y 49 DDKIVGYALDVSWPQVSIQYDKQHVCGGSILDPHWVLTAAHCFRKHTDVFNWKVRAG 108
b 183 EERIIGTQAEAGTGDWPPQVSLQNNVHCGGTLISNLWVLTAAHCFRSYSNPQQWTATFG 242
Y 109 SDKLGSFPLAVAKIIII--EFNPMYPKNDIALMKLQFPLTSGTVRPICLPFFDEEL 165
b 243 VSTIS--PRLRVRVRAILAHAEYNSI--TRNDIAVVQLDRPVTFTRNIHRVCLPAATQNI 299
Y 166 TPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMMKAGIPEG 225
b 300 MPDSVAVYVTGWSLTGYGNTVTN-LQQGEVRIVSSEVCNEPAGYGGSVLPGLMCGVRSRG 358
Y 226 GVDTCQDSDGGPLMYQSDQ--WHVVGIVSWGVCYGGPSTPGVYTKVSAYLNWI 276
b 359 AVDACQDSDGGPLVQEDTRRLWFVVGIVSWGVCYGLPNKPGVYTRVTAYRNWI 411
[1]
RESULT 13
Q8CDR0 PRELIMINARY; PRT; 455 AA.
D Q8CDR0
C Q8CDR0;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Transmembrane protease.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK029714; BAC26577.1; -
SQ SEQUENCE 455 AA; 49669 MW; BE22EB2E7503C74B CRC64;
Query Match 32.4%; Score 519; DB 11; Length 455;
Best Local Similarity 42.4%; Pred. NO. 2.9e-43;
Matches 101; Conservative 41; Mismatches 82; Indels 14; Gaps 6;
QY 51 KIVGGYALDVSWPQVSIQYDKQHVCGGSILDPHWVLTAAHCFRKHTDVFNWKVRA 107
Db 217 RIVGGQAVASGRWPQASVMLGSRHTCGASVLAPHWVVTAAHCFMYSFRL-SRLSSWRVHA 275
QY 108 GSDKLGs---FPSLAVAKIIIIIEFNPMYPKDN--DIALMKLQFPLTSGTVRPICLPFF 161
Db 276 GLVSHGAVRQHQGTWVEKIIP---HPLYSAQNHDYDVALQLRTPINFSDTVGAVCLPAK 332
QY 162 DEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQGEVTEKMMKAG 221
Db 333 EQHFPWGSQCWVSGWGHGTDPSHTSSDTLQDTMVPLLSTYLCNSSCMYSGALTHRMLCAG 392
QY 222 IPEGGVDTCCQDSDGGPLMYQS-DQWHVVGIVSWGVCYGGPSTPGVYTKVSAYLNWIYN 278
Db 393 YLDGRADACQDSDGGPLVCPSGDTHLVGVSWGRCAPNRPBGVYAKVAEFLDWIHD 450
[1]
RESULT 14
Q8VDV1 PRELIMINARY; PRT; 417 AA.
ID Q8VDV1
AC Q8VDV1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to airway trypsin-like protease.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; BC020151; AAH20151.1; -
DR HSSP; P00761; 1AN1.
DR MEROPS; S01.047; -
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000082; SEA domain.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF01390; SEA; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00200; SEA; 1.
DR SMART; SM00200; Tryp_Spc; 1.
DR PROSITE; PS50024; SEA; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.

W Hydrolase; Protease; Serine protease.
Q SEQUENCE 417 AA; 46254 MW; C7FACD21C06FEAEA CRC64;
Query Match 31.8%; Score 509.5; DB 11; Length 417;
Best Local Similarity 40.4%; Pred. No. 2.3e-42;
Matches 93; Conservative 42; Mismatches 92; Indels 3; Gaps 2;
Y 49 DDKIVGGYALDVSDFWQVSIQYDKQHVCGGSLDPHWLVLTAAHCFRKHDTDFVNWKVRA 108
b 183 EERIIGMGAEPGDWQVSLQLNNVHCGGALISNMWLVLTAAHCFKSPNPQYWTATFG 242
Y 109 SDKLGSFPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPA 168
b 243 VSTMSPLRVRRAILAHADGYSSVTRDNDIAVVQLDRSVAFSRNIHRVCLPAATQNIIPG 302
Y 169 TPLWIIGWGFTKQNGKMSDILLQASVOVIDSTRCNADDDAYQGEVTEKMCAGIPEGGVD 228
b 303 SVAYVTGWGSLTYGGNAVTN-LRQGEVRIISSEECNTPAGYSGSVLPGMLCAGMRSGAVD 361
Y 229 TCQGDGSGPLMYQSDQ--WHVVGVISWGYGCGGPSTPGVYTKVSAYLNWI 276
b 362 ACQGDGSGPLVQEDSRRLWFVVGIVSWGYQCGLPNKPGVYTRVTAYRNWI 411

RESULT 15
8VHK8 PRELIMINARY; PRT; 417 AA.
C Q8VHK8
T 01-MAR-2002 (TremBLrel. 20, Created)
T 01-MAR-2002 (TremBLrel. 20, Last sequence update)
T 01-MAR-2003 (TremBLrel. 23, Last annotation update)
E Airway trypsin-like protease.
S Mus musculus (Mouse).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
X NCBI_TaxID=10090;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=CZECH II; TISSUE=Trachea;
A Hansen I.A., Fassnacht M., Hammer F., Schammann M., Allolio B.;
T "Identification and localization of MAT - the murine homologue of the
human airway trypsin-like protease (HAT).";
L Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
C -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
R EMBL; AF448809; AAL47139.1; -.
R HSSP; P00761; 1AN1.
R MEROPS; S01.047; -.
R InterPro; IPR001314; Chymotrypsin.
R InterPro; IPR000082; SEA domain.
R InterPro; IPR001254; Ser_protease_Try.
R Pfam; PF01390; SEA; 1.
R Pfam; PF00089; trypsin; 1.
R PRINTS; PR00722; CHYMOTRYPSIN.
R SMART; SM00200; SEA; 1.
R SMART; SM00020; Tryp_SPC; 1.
R PROSITE; PS50024; SEA; 1.
R PROSITE; PS50240; TRYPSIN_DOM; 1.
R PROSITE; PS00134; TRYPSIN_HIS; 1.
R PROSITE; PS00135; TRYPSIN_SER; 1.
W Hydrolase; Protease; Serine protease.
Q SEQUENCE 417 AA; 46220 MW; 631ACF83C62DEB1E CRC64;

Query Match 31.8%; Score 509.5; DB 11; Length 417;
Best Local Similarity 40.4%; Pred. No. 2.3e-42;
Matches 93; Conservative 42; Mismatches 92; Indels 3; Gaps 2;
Y 49 DDKIVGGYALDVSDFWQVSIQYDKQHVCGGSLDPHWLVLTAAHCFRKHDTDFVNWKVRA 108
b 183 EERIIGMGAEPGDWQVSLQLNNVHCGGALISNMWLVLTAAHCFKSPNPQYWTATFG 242
Y 109 SDKLGSFPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPICLPFFDEELTPA 168
b 243 VSTMSPLRVRRAILAHADGYSSVTRDNDIAVVQLDRSVAFSRNIHRVCLPAATQNIIPG 302

QY 169 TPLWIIGWGFTKQNGKMSDILLQASVOVIDSTRCNADDDAYQGEVTEKMCAGIPEGGVD 228
Db 303 SVAYVTGWGSLTYGGNAVTN-LRQGEVRIISSEECNTPAGYSGSVLPGMLCAGMRSGAVD 361
QY 229 TCQGDGSGPLMYQSDQ--WHVVGVISWGYGCGGPSTPGVYTKVSAYLNWI 276
Db 362 ACQGDGSGPLVQEDSRRLWFVVGIVSWGYQCGLPNKPGVYTRVTAYRNWI 411

Search completed: November 23, 2003, 07:51:55
Job time : 44.7799 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

un on: November 23, 2003, 07:52:05 ; Search time 3687.55 Seconds
(without alignments)
3239.440 Million cell updates/sec

itle: US-09-607-745-9
erfect score: 1600
equence: 1 MDSKGSSQKSRLLLLLVSN.....LNWIYNVWKAELSRHHHHH 292

coring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

earched: 2888711 seqs, 2045481386 residues

otal number of hits satisfying chosen parameters: 5777422

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

ommand line parameters:

MODEL=frame+p2n.model -DEV=xlh
Q=/cgn2_1/USPTO spool/US09607745/runat_21112003_144345_22184/app_query.fasta_1.1038
DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
USER=US09607745 @CGN_1_1 5308 @runat_21112003_144345_22184 -NCPU=6 -ICPU=3
NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase : GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

Result No.	Score	Query Match %	Length	DB ID	Description
1	1266	79.1	1305	6	AX076190 Sequence
2	1266	79.1	1793	9	BC012752 Homo sapi
3	1266	79.1	2038	6	AR142620 Sequence
4	1266	79.1	2038	6	BD137129 Human pro
5	1266	79.1	2063	6	AX092380 Sequence
6	1266	79.1	2063	6	AX376262 Sequence
7	1266	79.1	2063	6	AX395214 Sequence
8	1266	79.1	2063	6	AX697206 Sequence
9	1266	79.1	2079	6	AR232520 Sequence
10	1266	79.1	2079	6	AX207967 Sequence
11	1266	79.1	2079	9	AF216312 Homo sapi
12	1266	79.1	2081	9	AF179224 Homo sapi
13	1266	79.1	2088	9	BC011703 Homo sapi
14	1264	79.0	699	6	E13203 Human cDNA
15	1253	78.3	1479	6	AX076192 Sequence
16	1248	78.0	2137	6	AX207899 Sequence
17	1050	65.6	2060	10	BC021368 Mus muscu
18	1050	65.6	2268	10	AY043240 Mus muscu
19	765	47.8	1130	6	AR256996 Sequence
20	754.5	47.2	1130	6	AR234337 Sequence
21	746	46.6	1169	6	AR219284 Sequence
22	740.5	46.3	1166	6	AR221273 Sequence
23	672.5	42.0	1052	6	AR219287 Sequence
24	645	40.3	1049	6	AR219286 Sequence
25	601.5	37.6	1103	6	AR219328 Sequence
26	593	37.1	1142	6	AR219285 Sequence
27	574	35.9	1037	6	AR219329 Sequence
28	546.5	34.2	1381	6	BD141852 Airway sp
29	539	33.7	2067	6	AX149581 Sequence
30	538	33.6	1555	6	BD141849 Airway sp
31	534.5	33.4	1538	6	BD141855 Airway sp
32	533	33.3	1078	6	AX190545 Sequence
33	533	33.3	1314	6	AX190543 Sequence
34	533	33.3	1314	6	AX190546 Sequence
35	533	33.3	1614	6	AX360089 Sequence
36	533	33.3	1689	6	AX149579 Sequence
37	533	33.3	1748	6	AX369323 Sequence
38	533	33.3	3222	9	BC018715 Homo sapi
39	533	33.3	3324	9	AB048797 Homo sapi
40	532	33.2	2237	9	AY190317 Homo sapi
41	528	33.0	2393	9	AB048796 Homo sapi
42	528	33.0	2992	10	BC042878 Mus muscu
43	527.5	33.0	1576	6	BD141854 Airway sp
44	527	32.9	1045	10	AF198087 Rattus no
45	526.5	32.9	1460	6	AR225450 Sequence

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ALIGNMENTS

```
X076190
LOCUS AX076190 1305 bp DNA linear PAT 06-FEB-2001
DEFINITION Sequence 1 from Patent WO0104141.
ACCESSION AX076190
VERSION AX076190.1 GI:12710815
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 Suendermann,B., Hofmann,U., Matzku,S. and Wilbert,O.
Seripancrin
Patent: WO 0104141-A 1 18-JAN-2001;
MERCK PATENT GmbH (DE)
FEATURES
Location/Qualifiers
1..1305
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1..>1305
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC28479.1"
/db_xref="GI:12710816"
/translation="MDPDSQPLNSLDVLPKPRIPMETFRKVGIPIIIALSLASI
IIVVLKIVLDKYFLCGQLHPIPRKQLCDGELDCPLGEDEHCVKSFPEGAVAL
RLSKDRSTLQVLDSATGNWFSACFDNFTEALAEATACRMGYSSKPTFRFAVEIGPDOL
DVVEITENSQELMRNNSGPGCLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWFWQV
SIOYDKOHVCGGSIIDPHWLTAHCFRKHDTDFVNWKVRAGSKLGSFPPLAVAKIIL
IEFNPMYKNDIALMKLQFPPLTFSGTVRPICLPFFDEELTPATPLWIIIGWFTKQNG
GKMSDILLQASVQVIDSTRNADDAVYQEVTEKMWKAGIPEGGVDTCCQDSSGGLPMVQ
SDQHWVGVISWVGCGGPGSTPGVYTKVSAVLNWIYNVWKAEL"
BASE COUNT 293 a 368 c 372 g 272 t
ORIGIN

Alignment Scores:
red. No.: 6.46e-122 Length: 1305
score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
B: 6 Gaps: 0

us-09-607-745-9 (1-292) x AX076190 (1-1305)
AY 51 LysileValGlyGlyTyrAlaLeuAspValAspSerTrpTrpGlnValSerIleGln 70
: : : : :
b 604 CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGGTCCAGTCCAG 663
AY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
: : : : :
b 664 TAGACAAACAGCACGCTCTGTGGAGGAGCATCTCTGGACCCCACTGGTCTCCAGGCA 723
AY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
: : : : :
b 724 GCCCACCTGCTTCAGAAACATACCGATGTGTTCAACTGGAAGTGCGGGCAGGCTCAGAC 783
AY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
: : : : :
b 784 AAACCTGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATCATTTGAATCAACCC 843
AY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
: : : : :
b 844 ATGTACCCCAAGAACAAATGACATCGCCCTCATGAAGCTGCAGTCCCACCTCTCTCA 903
AY 151 GlyThrValArgProLysCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
: : : : :
b 904 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 963
AY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
: : : : :
b 964 CTCTGGATCATTTGGATGGGCTTTACGAACGAGAATGGAGGGAAGATGTCTGCACATACTG 1023
```

```
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
: : : : :
Db 1024 CTGCAGGGCTCAGTCCAGTCAATGACAGCACACGGTGCAATGCACGATGGTACCAG 1083
QY 211 GlyGlnValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
: : : : :
Db 1084 GGGGAAGTCACCGAGAGATGATGTGTGCAGGCATCCCGGAAGGGGTGTGGACACCTGC 1143
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValGlyIle 250
: : : : :
Db 1144 CAGGTGACAGTGTGGGCCCTGATGTATCCAAATCTGACCACTGCGCATGTGTGGGCATC 1203
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
: : : : :
Db 1204 GTTAGCTGGGGCTATGGCTGGGGGGGGCCCGAGCACCCCGAGGATATACACCAAGGTCTCA 1263
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
: : : : :
Db 1264 GCCTATCTCAACTGGAATCTACAATGTCTGGAAGGCTGAGCTG 1305

RESULT 2
BC012752 1793 bp mRNA linear PRI 20-AUG-2001
LOCUS Homo sapiens, Similar to transmembrane protease, serine 4, clone
DEFINITION IMAGE:3623466, mRNA, partial cds.
ACCESSION BC012752
VERSION BC012752.1 GI:15215322
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1793)
REFERENCE Strausberg,R.
AUTHORS Direct Submission
TITLE Submitted (15-AUG-2001) National Institutes of Health, Mammalian
JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbiology.org
contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 24 Row: m Column: 19.
Location/Qualifiers
1..1793
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3623466"
/tissue type="pancreas, adenocarcinoma"
/clone_lib="NIH_MGC_39"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
<1..1219
/codon_start=2
/product="Similar to transmembrane protease, serine 4"
/protein_id="AAH12752.1"
/db_xref="GI:15215323"
/translation="ARGIIALLSLASIIIVVLIKVIDKYFLCGQLHFIPRKQL
CDGELDCPLGEDEHCVKSFPEGAVALRSLKDRSTLQVLDSATGNWFSACFDNFTEA
LAETACRMGYSSKPTFRFAVEIGPDQLDVVEITENSQELMRNNSGPGCLSGSLVSLH
```

CLACGKSLKTPRVVGGHEASVDSWPQVQSIQYDKQHVCGGSIIDPHVLTAAHCFRKH
TDVFNKVRAGSKLGSFPLSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVP
ICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQGEV
TEKMMGAGIPEGGVDTCCQDGGGGLMYQSDQHVHVIVSWGVCYCGPSTPGVYTKVSA
YLNWIYNVWKAEL"
BASE COUNT 436 a 501 c 486 g 370 t
ORIGIN

Alignment Scores:
Pred. No.: 9.65e-122 Length: 1793
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
Gaps: 9

US-09-607-745-9 (1-292) x BC012752 (1-1793)

Y 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
b 515 CGTGTGGTGGTGGGAGGAGCCCTCTGTGGATTCTTGGCCTTGGCAGGTGAGCATCCAG 574
Y 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
b 575 TACGACAAACAGCACAGCTCTGTGGAGGGAGCATCTCGACCCCCACCTGGGTCTCACGGCA 634
Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
b 635 GCCCACTGCTTCAGGAAACATACCGATGTGTTCACTGGAAGGTGCGGGCAGGCTCAGAC 694
Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
b 695 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTTGAATTCACACCCC 754
Y 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
b 755 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCCACTCACTTTCTCA 814
Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
b 815 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGTCACTCCAGCCACCCCA 874
Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
b 875 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGAGATGTGTGACATACCTG 934
Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
b 935 CTGCAGGCTCAGTCCAGGTCAATGACAGCACACGGTGCAATGCAGACGATCGTACCAG 994
Y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
b 995 GGGGAAGTCACCGAGAAGATGATGTGCAGGCATCCCGAAGGGGCTGGACACCTGC 1054
Y 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
b 1055 CAGGGTGACAGTGGTGGCCCTGATGTACCAATCTGACCAGTGGCATGTGTGGGCATC 1114
Y 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
b 1115 GTTAGCTGGGCTATGGCTGCGGGGCGCCGAGACCCAGGAGTATACCAAGGTCTCA 1174
Y 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
b 1175 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1216

RESULT 3
R142620
OCUS
DEFINITION Sequence 18 from patent US 6203979.
ACCESSION AR142620
VERSION AR142620.1 GI:15103906
KEYWORDS

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2038)
AUTHORS Bandman,O., Hillman,J.L., Yue,H., Guegler,K.J., Corley,N.C.,
Tang,Y.Tom. and Shah,P.
TITLE Human protease molecules
JOURNAL Patent: US 6203979-A 18 20-MAR-2001;
FEATURES Location/Qualifiers
source i. .2038
BASE COUNT 462 a 591 c 569 g 416 t
ORIGIN

Alignment Scores:
Pred. No.: 1.13e-121 Length: 2038
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
Gaps: 6

US-09-607-745-9 (1-292) x AR142620 (1-2038)

QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
Db 803 CGTGTGGTGGTGGGAGGAGCCCTCTGTGGATTCTTGGCCTTGGCAGGTGAGCATCCAG 862
QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
Db 863 TACGACAAACAGCACAGCTCTGTGGAGGGAGCATCTCGACCCCCACCTGGGTCTCACGGCA 922
QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
Db 923 GCCCACTGCTTCAGGAAACATACCGATGTGTTCACTGGAAGGTGCGGGCAGGCTCAGAC 982
QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
Db 983 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTTGAATTCACACCCC 1042
QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
Db 1043 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCCACTCACTTTCTCA 1102
QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
Db 1103 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGTCACTCCAGCCACCCCA 1162
QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
Db 1163 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTGTGACATACCTG 1222
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
Db 1223 CTGCAGGCTCAGTCCAGGTCAATTGACAGCACACCGTGCATGTCAGACGATCGGTACCAG 1282
QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
Db 1283 GGGGAAGTCACCGAGAAGATGATGTGCAGGCATCCCGAAGGGGCTGTGGACACCTGC 1342
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
Db 1343 CAGGGTGACAGTGGTGGCCCTGATGTACCAATCTGACCAGTGGCATGTGTGGTGGGCATC 1402
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
Db 1403 GTTAGCTGGGCTATGGCTGCGGGGCGCCGAGCACCCAGGAGTATACCAAGGTCTCA 1462
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1463 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1504

RESULT 4

BD137129
LOCUS BD137129 2038 bp DNA linear PAT 18-SEP-2002
DEFINITION Human protease molecule.
ACCESSION BD137129
VERSION BD137129.1 GI:23232074
KEYWORDS JP 2002508970-A/6.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2038)
AUTHORS Bandman,O., Hillman,J.L., Yue,H., Guegler,K.J., Corley,N.C.,
Tang,T.Y. and Shah,P.
TITLE Human protease molecule
JOURNAL Patent: JP 2002508970-A 6 26-MAR-2002;
INCYTE PHARMACEUTICALS INC
COMMENT OS Homo sapiens (human)
PN JP 2002508970-A/6
PD 26-MAR-2002
PF 12-JAN-1999 JP 2000540252
PR 16-JAN-1998 US 09/008271
PI OLGA BANDMAN, JENNIFER L. HILLMAN, HENRY YUE, KARL J. GUEGLER, NEIL
PI C CORLEY,
PI TOM Y. TANG, PURVI SHAH
PC C12N15/09, A61K38/46, C07K16/40, C12N1/19, C12N1/21, C12N5/10 PC
, C12N9/48, C12N9/64,
PC C12Q1/68, C12N15/00, A61K37/54, C12N5/00
CC 1337018
FH Key Location/Qualifiers
FT source 1..2038
FT /organism='Homo sapiens (human)'.
FEATURES
source Location/Qualifiers
1..2038
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 462 a 591 c 569 g 416 t
ORIGIN
Alignment Scores:
Pred. No.: 1.13e-121 Length: 2038
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 6 Gaps: 0
US-09-607-745-9 (1-292) x BD137129 (1-2038)
AY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
: : : : :
AB 803 CGTGTGGTGGTGGGAGGAGGCTCTGTGATCTTGGCCTTGGCAGGTCAGCATCCAG 862
AY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
: : : : :
AB 863 TACGACAAACAGCACGTCCTGTGGAGGAGCATCTCTGACCCCACTGGTCTCACGGCA 922
AY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
: : : : :
AB 923 GCCCACTGCTTCAGGAACATACCGATGTGTTCACTGGAAGGTGGGGCAGGCTCAGAC 982
AY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
: : : : :
AB 983 AAATGGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATATTGAATCAACCCC 1042
AY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
: : : : :
AB 1043 ATGTACCCCAAAGACAATGACATGCCCTCATGAAGTGCAGTTCCTCACTCTCTCA 1102
AY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
: : : : :
AB 1103 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1162

QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
: : : : :
Db 1163 CTCTGGATCATTTGGGCTTTACGAAGCAGATGGAGGAAATGTTGACATCTG 1222
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
: : : : :
Db 1223 CTGCAGCGTCACTCCAGGTCAATTGACAGCACACGGTGCATGACAGCATCGTACCAG 1282
QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
: : : : :
Db 1283 GGGGAAGTCACCGAGAAGATGATGTGCAGGCATCCCGAAGGGGTGTGGACACCTGC 1342
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
: : : : :
Db 1343 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGCATC 1402
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
: : : : :
Db 1403 GTTAGCTGGGCTATGGCTCGGGGGCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1462
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
: : : : :
Db 1463 GCCTATCTCACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1504
RESULT 5
AX092380
LOCUS AX092380 2063 bp DNA linear PAT 21-MAR-2001
DEFINITION Sequence 111 from Patent WO0116318.
ACCESSION AX092380
VERSION AX092380.1 GI:13444504
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Eaton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A.,
Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and
Wood,W.I.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0116318-A 111 08-MAR-2001;
Genentech, Inc. (US)
FEATURES
source Location/Qualifiers
1..2063
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 477 a 591 c 576 g 419 t
ORIGIN
Alignment Scores:
Pred. No.: 1.15e-121 Length: 2063
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 6 Gaps: 0
US-09-607-745-9 (1-292) x AX092380 (1-2063)
QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
: : : : :
Db 804 CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTTCTTGGCCTTGGCAGGTCAGCATCCAG 863
QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
: : : : :
Db 864 TACGACAAACAGCACGTCCTGTGGAGGAGCATCTCTGACCCCACTGGTCTCTCACGGCA 923
QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
: : : : :
Db 924 GCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGTGGGGCAGGCTCAGAC 983

Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
b 984 AAACGGGAGCTTCCCATCCCTGGCTGGCCAAAGATCATCATTTGAATCAACCCC 1043
Y 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
b 1044 ATGTACCCCAAGACATGACATCGCCCTCATGAAGTGCAGTCCCACTCACTTTCTCA 1103
Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAlaThrPro 170
b 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163
Y 171 LeuTyrIleIleGlyTyrGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu 190
b 1164 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGAAGATGTCTGACATACTG 1223
Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 210
b 1224 CTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCAATGCAGACGATGCGTACCAG 1283
Y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
b 1284 GGGGAAGTCACCGAGAAGATGATGTGCGAGGCATCCCGAAGGGGTGTGGACACCTGC 1343
Y 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValGlyIle 250
b 1344 CAGGGTGACAGTGGTGGCCCTGATGTACCAATCTGACCAGTGGCATGTTGGTGCAATC 1403
Y 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
b 1404 GTTAGCTGGGCTATGGCTGGGGCCCGAGCACCCAGAGATATACCAAGGTCTCA 1463
Y 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
b 1464 GCCTATCTCAACTGGATCTACATGTCTGGAAGGCTGAGCTG 1505
RESULT 6
X376262
OCUS
AX376262
DEFINITION Sequence 329 from Patent WO0168848.
CESSION AX376262
VERSION AX376262.1 GI:19170522
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Baker, K.P., Chen, J., Desnoyers, L., Goddard, A., Godowski, P.J.,
Gurney, A.L., Pan, J., Smith, V., Watanabe, C.K., Wood, W.I. and
Zhang, Z.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0168848-A 329 20-SEP-2001;
Genentech, Inc. (US)
FEATURES
source Location/Qualifiers
1. .2063
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ASE COUNT 477 a 591 c 576 g 419 t
RIGIN

Alignment Scores:
red. No.: 1.15e-121 Length: 2063
core: 1266.00 Matches: 228
percent Similarity: 98.29% Conservative: 2
est Local Similarity: 97.44% Mismatches: 4
very Match: 79.12% Indels: 0
B: 6 Gaps: 0
S-09-607-745-9 (1-292) x AX376262 (1-2063)

QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
Db :::: :
804 CGTGTGGTGGTGGGAGGAGGCTCTCTGTGATTTCTGGCCTTGGCAGGTCCAGTCCAG 863
QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
Db TACGACAAACAGCAGCTCTGTGGAGGAGCATCTGGACCCCACTGGGTCTCAGCGCA 923
QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
Db GCCCACTGCTTCAGGAACATACCGATGTGTTCACTGGAAGGTGGGGCAGGCTCAGAC 983
QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleLeuGluPheAsnPro 130
Db AAACGGGAGCTTCCCATCCCTGGCTGGCCAAAGATCATCATTTGAATCAACCCC 1043
QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
Db ATGTACCCCAAGACATGACATCGCCCTCATGAAGTGCAGTCCCACTCACTTTCTCA 1103
QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
Db GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163
QY 171 LeuTyrIleIleGlyTyrGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu 190
Db CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGAAGATGTCTGACATACTG 1223
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 210
Db CTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCAATGCAGACGATGCGTACCAG 1283
QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
Db GGGGAAGTCACCGAGAAGATGATGTGCGAGGCATCCCGAAGGGGTGTGGACACCTGC 1343
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValGlyIle 250
Db CAGGGTGACAGTGGTGGCCCTGATGTACCAATCTGACCAGTGGCATGTTGGTGCAATC 1403
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
Db GTTAGCTGGGCTATGGCTGGGGCCCGAGCACCCAGAGATATACCAAGGTCTCA 1463
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db GCCTATCTCAACTGGATCTACATGTCTGGAAGGCTGAGCTG 1505
RESULT 7
AX395214
LOCUS AX395214 2063 bp DNA linear PAT 18-MAY-2002
DEFINITION Sequence 2 from Patent WO0216429.
ACCESSION AX395214
VERSION AX395214.1 GI:21066245
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Goddard, A., Godowski, P.J., Gurney, A.L., Hillan, K.J., Polakis, P.,
Smith, V., Wood, W.I., Wu, T.D. and Zhang, Z.
TITLE Compositions and methods for the diagnosis and treatment of tumor
JOURNAL Patent: WO 0216429-A 2 28-FEB-2002;
Genentech, Inc. (US)
FEATURES
source Location/Qualifiers
1. .2063
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 477 a 591 c 576 g 419 t
ORIGIN

Alignment Scores:

red. No.: 1.15e-121 Length: 2063
core: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 6 Gaps: 0

S-09-607-745-9 (1-292) x AX395214 (1-2063)

Y 51 LysileValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
b 804 CGTGTGGTGGTGGGAGGAGCCCTCTGTGGATTCTTGGCCTTGGCAGTCAGCATCCAG 863
Y 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
b 864 TAGCACAAACAGCACGCTCTGTGGAGGAGCATCTGGACCCCACTGGGTCTCACGGCA 923
Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
b 924 GCCCACTGCTTCAGGAACATACCGATGTGTCAACTGGAAGGTGGCGCAGGCTCAGAC 983
Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
b 984 AAAGTGGGAGCTTCCCATCCCTGGCTGTGGCAAGATCATCATTAATTCACACCCC 1043
Y 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLysPheSer 150
b 1044 ATGTACCCCAAGACATGACATCGCCCTCATGAAGCTGCAGTTCCCACTCACTTTCTCA 1103
Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
b 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163
Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
b 1164 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACTG 1223
Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
b 1224 CTGCAGGCGTCAGTCCAGGTCATTGACAGCACAGGTCGCAATGCAGACGATCGGTACCAG 1283
Y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
b 1284 GGGGAAGTCACCGAGAAGATGATGTGTGACGGCATCCCGAAGGGGTGTGGACACCTGC 1343
Y 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
b 1344 CAGGGTGACAGTGGTGGCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGCATC 1403
Y 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
b 1404 GTTAGCTGGGCTATGGCTCGGGGGCCCGACACCCAGGAGTATACACCAAGGTCTCA 1463
Y 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
b 1464 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505

RESULT 8

AX697206 2063 bp DNA linear PAT 02-APR-2003
SEQUENCE 274 from Patent WO0078961.

AX697206

AX697206.1 GI:29498146

Homo sapiens (human)

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1

REFERENCE

AUTHORS

Ferrara, N., Stewart, T.A., Williams, P.M., Baker, K.P., Desnoyers, L.,

Eaton, D.L., Gao, W.Q., Pan, J., Botstein, D., Fong, S., Goddard, A.,

Godowski, P.J., Gurney, A.L., Smith, V., Tumas, D., Wood, W.I.,

Grimaldi, C.J., Hillan, K.J., Paoni, N.F., Roy, M.A. and Watanabe, C.K.
Secreted and transmembrane polypeptides and nucleic acids encoding
the same

Patent: WO 0078961-A 274 28-DEC-2000;

Genentech Inc. (US)

FEATURES Location/Qualifiers

source

1. 2063

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

BASE COUNT 477 a 591 c 576 g 419 t

ORIGIN

Alignment Scores:

Pred. No.: 1.15e-121 Length: 2063

Score: 1266.00 Matches: 228

Percent Similarity: 98.29% Conservative: 2

Best Local Similarity: 97.44% Mismatches: 4

Query Match: 79.12% Indels: 0

DB: 6 Gaps: 0

US-09-607-745-9 (1-292) x AX697206 (1-2063)

QY 51 LysileValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70

Db 804 CGTGTGGTGGTGGGAGGAGCCCTCTGTGGATTCTTGGCCTTGGCAGTCAGCATCCAG 863

QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90

Db 864 TAGCACAAACAGCACGCTCTGTGGAGGAGCATCTGGACCCCACTGGGTCTCACGGCA 923

QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110

Db 924 GCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAGGTGGGGCAGGCTCAGAC 983

QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130

Db 984 AAAGTGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTAATTCACACCC 1043

QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150

Db 1044 ATGTACCCCAAGACATGACATCGCCCTCATGAAGTCGCAATGCAGACGATCGGTACCAG 1103

QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170

Db 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163

QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190

Db 1164 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACTG 1223

QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210

Db 1224 CTGCAGGCGTCAGTCCAGGTCATTGACAGCACACCGTGCATGCAGACGATCGGTACCAG 1283

QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230

Db 1284 GGGGAAGTCACCGAGAAGATGATGTGTGTCAGGCTATCCCGAAGGGGTGTGGACACCTGC 1343

QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250

Db 1344 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGCATC 1403

QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270

Db 1404 GTTAGCTGGGCTATGGCTCGGGGGCCCGACACCCAGGAGTATACACCAAGGTCTCA 1463

QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284

Db 1464 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505

RESULT 9

AR232520

OCUS AR232520 2079 bp DNA linear PAT 20-DEC-2002
EFINITION Sequence 1 from patent US 6455668.
CESSION AR232520
ERSION AR232520.1 GI:27274657
YWORDS
OURCE Unknown.
ORGANISM Unknown.
Unclassified.
EFERENCE 1 (bases 1 to 2079)
AUTHORS Mack,D., Gish,K.C. and Wilson,K.E.
TITLE Methods of diagnosing colorectal cancer, compositions, and methods
of screening for colorectal cancer modulators
JOURNAL Patent: US 6455668-A 1 24-SEP-2002;
EATURES Location/Qualifiers
source 1. .2079
/organism="unknown"
ASE COUNT 489 a 594 c 575 g 421 t
RIGIN
lignment Scores:
red. No.: 1.16e-121 Length: 2079
core: 1266.00 Matches: 228
ercent Similarity: 98.29% Conservatative: 2
est Local Similarity: 97.44% Mismatches: 4
uery Match: 79.12% Indels: 0
B: 6 Gaps: 0
S-09-607-745-9 (1-292) x AR232520 (1-2079)
Y 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
b 818 CGTGTGGTGGTGGGAGGAGGCCTCTGTGGATCTTGGCCTTGGCAGGTCAGCATCCAG 877
Y 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
b 878 TAGACAAACAGCACGCTGTGTGGAGGAGCATCTGGACCCCACTGGTCTCACGGCA 937
Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
b 938 GCCCACTGCTTCAGGAACATACCGATGTGTCAACTGGAAGTGGGGCAGGCTCAGAC 997
Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
b 998 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCAAGATCATCATATGATGATCAACCCC 1057
Y 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
b 1058 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGAGTCCCACTCACTTTCTCA 1117
Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
b 1118 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCCCA 1177
Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
b 1178 CTCTGGATCATTGGATGGGGCTTTACGACAGCAATGGAGGAAGATGTCTGACATACTG 1237
Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 210
b 1238 CTGCAGCGTCAGTCCAGGTCAATTGACAGCACACGGTGTCAATGAGACGATCGGTACCG 1297
Y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
b 1298 GGGGAAGTCACCGAGAAGATGATGTGTGAGGCATCCCGGAAGGGGTGTGGACACCTGC 1357
Y 231 GlnGlyAspSerGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
b 1358 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAAGTGGCATGTGTGGGCATC 1417
Y 251 ValSerTrpGlyTyrGlyCysGlyProSerThrProGlyValTyrThrLysValSer 270
b 1418 GTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCCAAGGAGTATACCAAGGTTCTCA 1477

QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
|||||
Db 1478 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1519
RESULT 10
AX207967
LOCUS AX207967 2079 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 71 from Patent WO0157194.
ACCESSION AX207967
VERSION AX207967.1 GI:15422563
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Madison,E.L., Ong,E.O. and Yeh,J.C.
TITLE Nucleic acid molecules encoding transmembrane serine proteases, the
encoded proteins and methods based thereon
JOURNAL Patent: WO 0157194-A 71 09-AUG-2001;
CORVAS INTERNATIONAL, INC. (US)
FEATURES
source Location/Qualifiers
1. .2079
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
251..1522
/note="Nucleotide sequence encoding transmembrane
protease, serine 4 (TMPRSS4)"
/codon_start=1
/protein_id="CAC60389.1"
/db_xref="GI:15422564"
/translation="MSNPCANPVSPWRPSESVGIPITIALLSLASIIIVVVLIKVILD
KYVFLCGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSFPEGPAAVRLSKDRSLQVL
DSATGNWFSACFDNFTEALAEATACRQMGYSKPTFRAVEIGPDQDLDDVVEITENSQEL
RMRNSSGPCLSGLVSLHCLACGKSLKTPRVVGSEASVDSWPQVSIQYDKQHVCGG
SILDPHWLTAACHFRKHTDVFNWKVRAGSDKLGSPPSLAVAKIIIEFNPMYPKDND
IALMKLOFPLTFSGTVRPICLPFDEELTPATPLNIIGWFTKQNGGKMSDILLQASV
QVIDSTRCNADDAYQGEVTEKMCAGIPEGGVDTQGDSSGGLMYQSDQWHVVGIVSW
GYCGGPGSTPGVYTKVSAYLNWIYNVWKAEL"
BASE COUNT 489 a 594 c 575 g 421 t
ORIGIN
Alignment Scores:
Pred. No.: 1.16e-121 Length: 2079
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservatative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 6 Gaps: 0
US-09-607-745-9 (1-292) x AX207967 (1-2079)
QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
b 818 CGTGTGGTGGTGGGAGGAGGCCTCTGTGGATCTTGGCCTTGGCAGGTCAGCATCCAG 877
QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
b 878 TAGACAAACAGCACGCTGTGTGAGGGAGCATCTTGGACCCCACTGGTCTCACGGCA 937
QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
b 938 GCCCACTGCTTCAGGAACATACCGATGTGTCAACTGGAAGTGGGGCAGGCTCAGAC 997
QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
b 998 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCAAGATCATCATATGATGATCAACCCC 1057
QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
b 1058 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGAGTCCCACTCACTTTCTCA 1477

```
Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
|||||
b 1118 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1177
|||||
Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
|||||
b 1178 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGAGATGTTGACATACTG 1237
|||||
Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 210
|||||
b 1238 CTGCAGGCGTCAGTCCAGTCAATTGACAGCACACGGTGCATGCAGACGATCGTACCAG 1297
|||||
Y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
|||||
b 1298 GGGGAAGTCACCGAGAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGACACCTGC 1357
|||||
Y 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
|||||
b 1358 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGTGGGCATC 1417
|||||
Y 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTrpThrLysValSer 270
|||||
b 1418 GTTAGCTGGGGCTATGGCTCGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1477
|||||
Y 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
|||||
b 1478 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1519
|||||
RESULT 11
F216312
OCUS
DEFINITION Homo sapiens type II membrane serine protease mRNA, complete cds.
ACCESSION AF216312
VERSION AF216312.1 GI:6911218
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Smeekens,S.S., Lorimer,D.D., Wang,E., Hou,J. and Linnevers,C.
MT-SP2, a novel type II membrane serine protease expressed in
trachea, colon, and small intestine: identification, cloning, and
chromosomal localization
Unpublished
JOURNAL
DEFINITION 2 (bases 1 to 2079)
AUTHORS Smeekens,S.S., Lorimer,D.D., Wang,E., Hou,J. and Linnevers,C.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-1999) Axys Pharmaceuticals, Inc, 180 Kimball Way,
South San Francisco, CA 94080, USA
FEATURES
source
1..2079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
251..1522
/note="MT-SP2"
/codon_start=1
/product="type II membrane serine protease"
/protein_id="AAF31436.1"
/db_xref="GI:6911219"
/translation="MSNPCANPVSPRPSESVGIPITIALSLASIIIVVLKIVILD
KYATLCGQLHFIPIRKQLCDGELDCPLGEDEHCVKSPFPGPAVAVRLSKDRSTLQVL
DSYFLGNWESACFDNFTEALAEACFQMGYSKPTFRAVEIGPDQDLVDVEITENSQEL
RMRNSSGCLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGG
SILDPHWLTAACFRKHTDVFNNKVRAGSDKLSGFSLAVAKIIIEFNPMYPKDN
IALMKLQPLTFSTGVRPICLPFDEELTPATPLWIIWGFTKQNGGKMSDILLQASV
QVIDTRCNADDAAYQGEVTEKMCAGIPEGGVDTCCQSGSGLPMYQSDQWHVVGVISW
GYCGGSPSTPGVYTKVSAYLNIYNVWKAEL"
BASE COUNT 489 a 594 c 575 g 421 t
ORIGIN
...
Alignment Scores:
```

```
Pred. No.: 1.16e-121 Length: 2079
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 9 Gaps: 0
US-09-607-745-9 (1-292) x AF216312 (1-2079)
QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpTrpGlnValSerIleGln 70
|||||
Db 818 CGTGTGGTGGTGGGAGGAGGCTCTGTGGATCTTGGCCTTGGCAGGTTCAGCATCCAG 877
|||||
QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
|||||
Db 878 TACGACAAACACAGCACAGTCTGTGGAGGAGCATCTTGACCCCACTGGTCTCTCACGGCA 937
|||||
QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
|||||
Db 938 GCCCACTGCTTCAGGAACAATACCATGTTGTTCAACTGGAAAGGTGGGGCAGGCTCAGAC 997
|||||
QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
|||||
Db 998 AAACCTGGGAGCTTCCCATCCCTGGCTGGCCAAAGATCATCATCATTTCAACCCC 1057
|||||
QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
|||||
Db 1058 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGTGCAGTCCCACTCACTTTCTCA 1117
|||||
QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
|||||
Db 1118 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGTCACTCCAGCCACCCCA 1177
|||||
QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
|||||
Db 1178 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAGATGTCTGACATACTG 1237
|||||
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 210
|||||
Db 1238 CTGCAGGCGTCAGTCCAGTCAATTGACAGCACACGGTGCATGCAGACGATCGTACCAG 1297
|||||
QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
|||||
Db 1298 GGGGAAGTCACCGAGAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGACACCTGC 1357
|||||
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
|||||
Db 1358 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAATCTGACAGTGGCATGTGTGGGCATC 1417
|||||
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTrpThrLysValSer 270
|||||
Db 1418 GTTAGCTGGGGCTATGGCTCGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1477
|||||
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
|||||
Db 1478 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1519
|||||
RESULT 12
AF179224
LOCUS
DEFINITION Homo sapiens transmembrane serine protease 3 (TMPRSS3) mRNA,
complete cds.
ACCESSION AF179224
VERSION AF179224.1 GI:8347148
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2081)
Wallrapp,C., Hahnel,S., Muller-Pillasch,F., Burghardt,B.,
Iwamura,T., Ruthenburger,M., Lerch,M.M., Adler,G. and Gress,T.M.
A novel transmembrane serine protease (TMPRSS3) overexpressed in
```

pancreatic cancer
Cancer Res. 60 (10), 2602-2606 (2000)
20283276
PUBMED 10825129
2 (bases 1 to 2081)
Wallrapp,C. and Gress,T.M.
Direct Submission
Submitted (20-AUG-1999) Internal Medicine I, University of Ulm,
Robert-Koch-Street 8, Ulm, Baden-Wuerttemberg 89081, Germany
Location/Qualifiers
1. .2081
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q23.3"
/tissue type="pancreatic carcinoma"
1. .2081
/gene="TMPRSS3"
215. .1528
/gene="TMPRSS3"
/codon_start=1
/product="transmembrane serine protease 3"
/protein_id="AAF74526.1"
/db_xref="GI:8347149"
/translation="MLQDPDSQPLNSLDVPLRKPRIPIPETFRKVGIPILIALLSLA
SIIIVVLKIVLDKYFLCGQPLHFIIPKQLCDGELDCPLGEDEHCVKSFPEGPVAV
AVRLSKDRSTLQVLDSATGNWFSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQ
DLDVVEITENSQELNRNNSGPFCLSGSLVSLHCLACGSKLKTIPRVVGGEEASVDSWPW
QVSIQYDKQHVCGGSILDPHVLTAAHCFRKHDTDFNVKVRAGSKLGSFSPSLAVAKI
IIEFNPMYPKNDIALMKLQFPPLTFSGTVRPICLPFFDEELTPATPLWIIIGWGFTKQ
NGKMSDILLQASQVIDSTRCNADDAAYQGEVTEKMWKAGIPEGGVDTCCQGDSSGGPLM
YQSDQHWVGIWSGWYCGGPGSTPGVYTKVSAYLNIWYVWKAEI"
215. .307
/gene="TMPRSS3"
/note="Region: cytoplasmic domain"
308. .373
/gene="TMPRSS3"
/note="Region: transmembrane domain"
374. .1525
/gene="TMPRSS3"
/note="Region: extracellular domain"
602. .604
/gene="TMPRSS3"
/note="glycosylation site"
746. .748
/gene="TMPRSS3"
/note="glycosylation site"
order(800. .802,1142. .1144)
/gene="TMPRSS3"
/note="disulfide bond"
order(902. .904,950. .952)
/gene="TMPRSS3"
/note="disulfide bond"
947. .949
/gene="TMPRSS3"
/note="active site"
1082. .1084
/gene="TMPRSS3"
/note="active site"
order(1280. .1282,1328. .1330)
/gene="TMPRSS3"
/note="disulfide bond"
1373. .1375
/gene="TMPRSS3"
/note="active site"
484 a 597 c 576 g 424 t
BASE COUNT
RIGIN
alignent Scores:
red. No.: 1.17e-121 Length: 2081
core: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2

Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 9 Gaps: 0
US-09-607-745-9 (1-292) x AF179224 (1-2081)
QY 51 LysileValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
Db 824 CGTGTGGTGGTGGGAGGAGCCTCTGTGGATTCTTGGCCTTGGCAGGTGAGATCCAG 883
QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
Db 884 TACGACAAACAGCACGCTCTGTGGAGGAGCATCTCTGACCCCACTGGGTCTCAGGCA 943
QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
Db 944 GCCCACTGCTTCAGGAACATACCGATGTGTTCACTGGAAGGTGCGGGCAGGCTCAGAC 1003
QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
Db 1004 AAATGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATATTGAATCAACCCC 1063
QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
Db 1064 ATGTACCCCAAGACATGACATCGCCCTCATGAAGCTGCAGTTCCTCCACTCACTTCTCA 1123
QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
Db 1124 GGACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1183
QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
Db 1184 CTCTGGATCATTTGGATGGGCTTTACGAAGCAAGATGGAGGAAGATGCTCTGACATCTG 1243
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
Db 1244 CTGCAGGCTCAGTCCAGGTCAATTGACAGCACCGGTGCAATGCAGACGATCGGTACCAG 1303
QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
Db 1304 GGGGAAGTCACCGAAGATGATGTGTGACAGGATCCCGAAGGGGTGTGGACACCTGC 1363
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
Db 1364 CAGGTGACAGTGTGGGCCCTGTATGATACCAATCTGACCAAGTGGCATGTGGTGGCATC 1423
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
Db 1424 GTTAGCTGGGCTATGGCTCGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1483
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1484 GCCTATCTCACTGGATCTACATGTCTGTGGAAGGCTGAGCTG 1525
RESULT 13
BC011703
LOCUS BC011703 2088 bp mRNA linear PRI 02-AUG-2001
DEFINITION Homo sapiens, Similar to mosaic serine protease, clone MGC:19490
IMAGE:3610695, mRNA, complete cds.
ACCESSION BC011703
VERSION BC011703.1 GI:15079794
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2088)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nhgri.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 27 Row: f Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 8347148.

FEATURES
source

Location/Qualifiers
1..2088
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:19490 IMAGE:3610695"
/tissue_type="Pancreas, adenocarcinoma"
/clone_lib="NIH MGC 39"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
217..1530
/codon_start=1
/product="Similar to mosaic serine protease"
/protein_id="AAH11703.1"
/db_xref="GI:15079795"
/translation="MLQDPDSQPLNSLDVKPLRPIMETFRKVGIPILLLSLA
SIIIVVILKIVLDKYFLCGPLHPIPRKQLCDGELDCPLGDEEHCVKSPFEGPAV
AVRLSKDSTLQVLDSATGNWFSACFDNFTALAEATQMGYSKPTFRVIEIGPDQ
DLDVVEITENSQELMRNSSGCLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPW
QVSIQDKQHVCGGSILDPHWLTAACFRKHTDVFNNKVRAGSDKLGSPFSLAVAKI
IIIEFNMPKDNIDIALMKLQPLTFSGTVRPICLPFFDEHLETPATPLWIIGWFTKQ
NGKMSDILLQASQVIDSTRNADDAAYQGEVTEKVMCAGIPEGGVDTCCQGSGGPLM
YQSDWHVVGVISWVGCGGSPTPGVYTKVSAYLNIWYNWKAEL"
492 a 597 c 579 g 420 t

CDS

BASE COUNT
RIGIN

Alignment Scores:

red. No.: 1.17e-121 Length: 2088
score: 1266.00 Matches: 228
percent Similarity: 98.29% Conservative: 2
best Local Similarity: 97.44% Mismatches: 4
query Match: 79.12% Indels: 0
Gaps: 9

US-09-607-745-9 (1-292) x BC011703 (1-2088)

Y 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
b 826 CGTGTGGTGGTGGGAGGAGGCCTCTGTGGATTCTTGGCCTTGGCAGGTCAGCATCCAG 885
Y 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
b 886 TACGACAAACAGCACCTCTGTGGAGGGAGCATCTCTGGACCCCTCTGGTCTCTCACGGCA 945
Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
b 946 GCCCACTGCTTCAGAAACATACCGATGTTCACTGGAAGTGCAGGTCAGGCTCAGAC 1005
Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130

Db 1006 AAACGGGGAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATCATGAATTCACACCC 1065
Qy 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
Db 1066 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAAGTCCCACTCACTTTCTCA 1125
Qy 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
Db 1126 GGCACAGTCAGGCCCATCTGTCTGCCCTCTTTTGATGAGGAGCTCACTCCAGCCACCCCA 1185
Qy 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
Db 1186 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGGAGATGTCTGACATCTG 1245
Qy 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 210
Db 1246 CTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGCATGACAGCATGCGTACCAG 1305
Qy 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
Db 1306 GGGGAAGTCACCGAGAAGATGATGTGTCAGGCATCCCGAAGGGGTGTGGACACCTGC 1365
Qy 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
Db 1366 CAGGTTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAATGTCAGTGGCATGTGGTGGGCATC 1425
Qy 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
Db 1426 GTTAGCTGGGCTATGGCTGCGGGGCCCGAGCACCCCGAGAGTATACACCAAGGTCTCA 1485
Qy 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1486 GCCTATCTCACTGGATCTACATGTCTGAAGGCTGAGCTG 1527

RESULT 14

E13203
LOCUS Human cDNA encoding a serine protease. 699 bp DNA linear PAT 27-APR-1998
DEFINITION Human cDNA encoding a serine protease.
ACCESSION E13203
VERSION E13203.1 GI:3252008
KEYWORDS JP 1997149790-A/4.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 699)
AUTHORS Tsuruoka, N., Yamashiro, K., Tsujimoto, M. and Yamaguchi, M.
TITLE NEW SERINE PROTEASE
JOURNAL Patent: JP 1997149790-A 4 10-JUN-1997;
SUNTORY LTD
COMMENT OS Homo sapiens (human)
PN JP 1997149790-A/4
PD 10-JUN-1997
PF 24-JUL-1996 JP 1996212196
PR 29-SEP-1995 JP 95P 275105
PI TSURUOKA NOBUO, YAMASHIRO KYOKO, TSUJIMOTO MASAFUMI, PI
YAMAGUCHI NARE
PC C12N15/09, C07H21/04, C07K14/47, C12N1/21, C12N5/10, PC
C12N9/52//A61K38/46,
PC (C12N15/09, C12R1:91), (C12N1/21, C12R1:19), (C12N9/52, C12R1:19),
PC (C12N9/52,
PC C12R1:91);
CC strandedness: Double;
CC topology: Linear;
FH key Location/Qualifiers
FH source 1..699
FT /organism='Homo sapiens'
FT /cell_type='colon cancer-derived cell' FT
FT /cell_line='COLO 201'
FT /clone='SP60'
FT mat_peptide 1..699

FEATURES

source

Location/Qualifiers

1..699

/product='Serine protease'

/organism='Homo sapiens'

/mol_type='genomic DNA'

/db_xref='taxon:9606'

157 a 190 c 209 g 143 t

BASE COUNT

157 a 190 c 209 g 143 t

ALIGNMENT

SCORE

PERCENT SIMILARITY

BEST LOCAL SIMILARITY

QUERY MATCH

DB

US-09-607-745-9 (1-292) x E13203 (1-699)

Y 52 ILeValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGlnTyr 71

b 1 GTGGTGGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGGTCCAGTCCAGTAC 60

Y 72 AspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValIleThrAlaAla 91

b 61 GACAAACAGCACGCTCTGTGGAGGAGCATCTCTGGACCCCACTGGGTCTCTACGGCAGCC 120

Y 92 HisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAspLys 111

b 121 CACTGCTTCAGGAACATACCATGCTGTTCACACTGGAAGGTGGGGCAGGCTCAGACAAA 180

Y 112 LeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnProMet 131

b 181 CTGGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATTCATTGAATCAACCCCATG 240

Y 132 TyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGly 151

b 241 TACCCCAAGACAAATGATGATGCGCTCTATGAGTGCAGTTCCTCCACTCACTTCTCAGGC 300

Y 152 ThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrProLeu 171

b 301 ACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCATC 360

Y 172 TrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeuLeu 191

b 361 TGGATCATTTGGATGGGCTTACGAAGCAGATGGAGGGAAGATGCTGACATCTGCTG 420

Y 192 GlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGlnGly 211

b 421 CAGGCGTCAGTCCAGTCAATTGACAGCACACGGTGCAATGCAGATCGGTACCGAGGG 480

Y 212 GluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCysGln 231

b 481 GAAGTCACCGAGAAGATGATGTGTGAGGCAATCCCGAAGGGGTGTGGACACCTGCCAG 540

Y 232 GlyAspSerGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIleVal 251

b 541 GGTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCAAGTGGCATGCTGGTGGCATCGTT 600

Y 252 SerTrpGlyTyrGlyCysGlyProSerThrProGlyValTyrThrLysValSerAla 271

b 601 AGCTGGGCTATGGTCTGGGGGGCCCCGAGCACCCAGGAGTATACACCAAGGTCTCAGCC 660

Y 272 TyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284

b 661 TATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 699

OCUS

AX076192

Sequence 3 from Patent WO0104141.

CCESION

AX076192

VERSION

AX076192.1

GI:12710817

1479 bp

DNA

linear

PAT 06-FEB-2001

RESULT 15

AX076192

OCUS

Sequence 3 from Patent WO0104141.

CCESION

AX076192

VERSION

AX076192.1

GI:12710817

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..1479

/organism='Homo sapiens'

/mol_type='genomic DNA'

/db_xref='taxon:9606'

1..1479

/note='unnamed protein product'

/codon_start=1

/protein_id='CAC28480.1'

/db_xref='GI:12710818'

/translation='MDPDSQPLNSLDVKPLRPKPRIPMETFRKVGIPITIIALLSLASI

IIIVVLIKVIIDKYYFLCGQLHFIIPRKQLCDGELDCPLGEDEHCVKSFPEGPAVAV

RLSKDSTLQVLDSATGNWFSACFDNFTEALAEACQMGYSKPTFRFAVEIGPDODL

DVVEITENSQELMRNSSGCLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPWQV

SIQDKQHVCGGSILDPHWLTAACFRKHTDVFENWVKRAGSKLGSFSLAVAKII

IEFNPYKDNIDIALMKLPPLTFSGTRVPICLPFDDELTPATPLWIIIGWFTKQNG

GKMSDILLQASQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGVDTCQSDSGGGLMYQ

SDQHWVGVIVSWGYGCGGSETPGVYTKVSAYLNWYINWVKDRTIQRSNPSPTGLVIQ

QPAVPLMELRSARHTRLTTCVCFLSFPALQPLSPICFPFPWD"

BASE COUNT

329 a 425 c 407 g 318 t

ORIGIN

Alignment Scores:

Pred. No.: 1.7e-120

Score: 1253.00

Percent Similarity: 98.27%

Best Local Similarity: 97.40%

Query Match: 78.31%

DB: 6

US-09-607-745-9 (1-292) x AX076192 (1-1479)

QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70

Db 604 CGTGTGGTGGTGGGAGGAGGCGCTCTGTGGATTCTTGGCCTTGGCAGGTCCAGTCCAG 663

QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90

Db 664 TAGCACAAACAGCACGCTCTGTGGAGGAGGAGCATCTTGGACCCCACTGGTCTCTCACGGCA 723

QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110

Db 724 GCCCACTGCTTCAGGAACATACATACCATGTGTTCACACTGGAAGGTGCGGCGAGGCTCAGAC 783

QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130

Db 784 AAAGTGGGAGGCTTCCCATCCCTGGCTGTGGCCAGATCATCATTCATGAATTCACACCC 843

QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150

Db 844 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGTGCAGTTCACCACTCACTTTCTCA 903

QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170

Db 904 GGCACAGTCAGGCCCATCTGTGTGCCCTTCTTTGTATGAGGAGGCTCACTCCAGCCACCCCA 963

QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190

Db 964 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGAGGAGGAGATGTCTGACATACTG 1023

QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 210

Db 1024 CTGCAGGCGTCAGTCCAGGTTCATTGACAGCACACCGGTGCAATGCAGACGATGCGTACCAG 1083

```
Y      211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
b      1084 GGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGGAAGGGGGTGTGGACACCTGC 1143

Y      231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
b      1144 CAGGGTGACAGTGTGGGCCCCCTGATGTACCAATCTGACCAGTGGCATGTGTGGGCATC 1203

Y      251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
b      1204 GTTAGCTGGGGCTATGGCTGCGGGGCCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1263

Y      271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLys 281
b      1264 GCCTATCTCAACTGGATCTACAATGTCTTGGGAAG 1296
```

earch completed: November 23, 2003, 10:40:15
ob time : 3705.55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

un on: November 23, 2003, 07:49:55 ; Search time 263.081 Seconds
(without alignments)
2996.170 Million cell updates/sec

itle: US-09-607-745-9

erfect score: 1600

equence: 1 MDSKSSQSRLLLLLVSN.....LNWIYNVWKAELSRHHHHH 292

coring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

earched: 2552756 seqs, 1349719017 residues

otal number of hits satisfying chosen parameters: 5105512

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

ommand line parameters:

MODEL=frame+_p2n.model -DEV=xlh

Q=/cgn2_1/USPTO_spool/US09607745/runat_21112003_144344_22176/app_query.fasta_1.1038

DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0

LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi

LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15

MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

USER=US09607745@cgn_1_1_401@runat_21112003_144344_22176 -NCPU=6 -ICPU=3

NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG

DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase : N_Geneseq_19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1593	99.6	1189	24	AAI64290	Protease D-G catay
2	1266	79.1	1305	22	AAD02556	Human seripancrin
3	1266	79.1	1314	25	ABX76354	Lung cancer-associ
4	1266	79.1	1854	22	AAH99574	Human protein enco
5	1266	79.1	2038	20	AAX87154	Human protease HUP
6	1266	79.1	2063	21	AAA37099	Human PRO1570 (UNQ
7	1266	79.1	2063	22	AAS46089	Human DNA encoding
8	1266	79.1	2063	22	AAF92113	Human PRO1570 cDNA
9	1266	79.1	2063	22	AAF54396	DNA encoding prote
10	1266	79.1	2063	24	ABS74433	Human cDNA encodin
11	1266	79.1	2063	24	ABK11090	cDNA encoding tumo
12	1266	79.1	2063	25	ACA57847	Human PRO1570 cDNA
13	1266	79.1	2063	25	ACA58865	cDNA encoding huma
14	1266	79.1	2063	25	ACA60418	Novel human secret
15	1266	79.1	2063	25	ACA63428	cDNA encoding huma
16	1266	79.1	2063	25	ABX98317	Human cDNA encodin
17	1266	79.1	2063	25	ABX98819	Novel human secret
18	1266	79.1	2063	25	ACA05864	Human secreted/tra
19	1266	79.1	2063	25	ABX97908	Human PRO polynucl
20	1266	79.1	2063	25	ABX78692	Human PRO polynucl
21	1266	79.1	2063	25	ABX75705	Human cDNA encodin
22	1266	79.1	2063	25	ABX76910	Human PRO polynucl
23	1266	79.1	2063	25	ABX16750	Human cDNA encodin
24	1266	79.1	2079	22	AAD13169	Human transmembran
25	1266	79.1	2079	24	AAI72975	CJA8 cDNA. Homo s
26	1266	79.1	2079	25	ABS57763	cDNA encoding huma
27	1266	79.1	2081	24	AAI72976	CJA8 preferred cDN
28	1266	79.1	2165	24	ABZ35523	Human gene express
29	1266	79.1	2307	24	ABS76529	cDNA encoding huma
30	1264	79.0	699	18	AAI79127	Human serine prote
31	1259	78.7	2121	24	AAI64284	Human serine prote
32	1253	78.3	1479	22	AAD02557	Human seripancrin
33	1253	78.3	2070	21	AAZ90471	Cancer specific ge
34	1248	78.0	2137	22	AAD13114	Human membrane-ty
35	1246	77.9	1795	22	AAS26880	Human cDNA encodin
36	1230	76.9	1281	21	AAA97361	Human colorectal c
37	918	57.4	1084	22	AAH34953	Human colon cancer
38	806	50.4	1008	20	AAX04381	Human secreted pro
39	765	47.8	1130	22	AAS01366	Recombinant human
40	754.5	47.2	1130	22	AAF77000	Fusion gene of pro
41	746	46.6	1169	21	AAC87795	Activation constru
42	746	46.6	1169	22	AAF55267	Nucleotide sequenc
43	740.5	46.3	1166	22	AAD02991	Zymogen activation
44	672.5	42.0	1052	21	AAC87798	Activation constru
45	672.5	42.0	1052	22	AAF55270	Nucleotide sequenc

ALIGNMENTS

RESULT 1

AAI64290

ID AAI64290 standard; DNA; 1189 BP.

XX AAI64290;

AC AAI64290;

XX 07-MAY-2002 (first entry)

DT 07-MAY-2002 (first entry)

XX Protease D-G catalytic domain fusion gene construct encoding sequence.

DE Serine protease; D-G; human; zymogen; enzyme; cytostatic;

XX antinflammatory; dermatological; anticoagulation; cancer;

KW skin disorder; neuropathic pain; inflammatory disorder;

KW coagulation diathesis; thrombosis; laundry detergent; skin care;

KW gene therapy; gene; ds.

XX Homo sapiens.

OS

X		
H	Key	Location/Qualifiers
T	CDS	13..891
"		/*tag= a
T		/product= "protease D-G"
T		/transl_except= (pos: 640..642, aa: Gln)
T		13..93
T	sig_peptide	/*tag= b
T		/note= "Prolactin signal sequence"
T		94..888
T	mat_peptide	/*tag= c
T		/note= "Mature protease D-G"
T		

WO200202011-A1.

10-JAN-2002.

08-JUN-2001: 2001WO-US18568

30-JUN-2000: 2000US-0607745.

(ORTH) ORTHO-MCNEIL PHARM INC.

Darrow AL, Qi J, Andrade-Gordon P;

WPI; 2002-106601/14.

P-PSDB; AAG78578.

Nucleic acid encoding a serine protease called D-G protein which is useful for identifying modulators that are useful for treating a condition which is mediated by protease D-G, e.g. cancer, skin disorders, or neuropathic pain -

Claim 4; Fig 4B; 81pp; English.

The invention relates to an isolated and purified nucleic acid that encodes a serine protease called D-G protein. The activity of the protein of the invention may be described as cytostatic, antiinflammatory, dermatological and anticoagulation. The serine protease of the invention is a member of the trypsin/chymotrypsin-like (S1) serine protease family, which play an important role in processes such as digestion and regulatory amplification cascades through the proteolytic activation of inactive zymogen precursors. Protease D-G modulating compounds are useful for treating a condition which is mediated by protease D-G, e.g. cancer, skin disorders, neuropathic pain, inflammatory disorders, or coagulation diathesis/thrombosis. The polynucleotide encoding the protease is useful for identifying, detecting or isolating mutant forms of DNA molecules encoding the protease. The protease is useful for identifying modulators of the functional protease. The D-G protein can be used for formulation of compositions for laundry detergents and skin care products. Protease D-G gene therapy may be used to introduce protease D-G into the cells of target organisms. As the D-G protein is derived from a human, it is less likely to produce an allergic reaction in sensitive individuals when used in formulations for laundry detergents and skin care products. The current sequence represents the protease D-G catalytic domain in the zymogen activation construct encoding sequence.

Sequence 1189 BP; 297 A; 305 C; 307 G; 280 T; 0 other;

Alignment Scores:

pred. No.:	2.84e-154	Length:	1189
Core:	1593.00	Matches:	291
Percent Similarity:	99.66%	Conservative:	0
Best Local Similarity:	99.66%	Mismatches:	1
Query Match:	99.56%	Indels:	0
B:	24	Gaps:	0

IS-09-607-745-9 (1-292) x AAI64290 (1-1189)

1 MetAspSerLysGlySerSerGlnLysSerArgLeuLeuLeuValValSerAsn 20

b) 13 ATGGACAGCAAAGGTCGTCGAGAAATCCCGCCTGCTCCTGCTGCTGGTGGTCAAT 72

[illegible]

RESULT 2

AAD02556

ID AAD02556 standard; cDNA: 1305 BP.

AC AAD02556;

DT 02-MAY-2001 (first entry)

Human seripancrin cDNA.

Human; seripancrin; serine protease; chromosome 11q22-q23; therapy; arthritis; chronic obstructive pulmonary disorder; COPD; cancer; osteoporosis; aberrant wound healing; angiogenesis; diabetes; inflammatory disorder; stroke; cardiovascular disease; gene therapy vaccine; cytostatic; cerebroprotective; vulnerary; osteopathic; ss.

OS Homo sapiens.

X Detecting a lung cancer-associated transcript in a cell from a patient
T for treating lung cancer, by contacting a biological sample from the
T patient with a polynucleotide that exhibits increased or decreased
T expression in lung cancer -
X
S Claim 22; Page 353; 453pp; English.
X
C The invention relates to a method for detecting a lung cancer-associated
C transcript in a cell from a patient, comprising contacting a biological
C sample from the patient with a polynucleotide that selectively hybridises
C to a sequence that is at least 80 % identical to a gene that exhibits
C increased or decreased expression in lung cancer samples. Lung
C cancer-associated polynucleotides and polypeptides are used for
C identifying a compound that modulates a lung cancer-associated
C polypeptide, for inhibiting proliferation of a lung cancer-associated
C cell to treat lung cancer in a patient and for treating a mammal having
C lung cancer by administering a modulatory compound identified. The
C methods are useful for treating lung cancer, such as small cell lung
C cancer, non-small cell lung cancer or other benign or precancerous
C lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive
C pulmonary disease, fibrosis, hypersensitivity pneumonitis, interstitial
C pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides
C and polypeptides are useful for diagnostic purposes and as targets for
C screening for therapeutic compounds that modulate lung cancer, such as
C antibodies. Sequences ABX76124-ABX76474 represent lung cancer-associated
C polynucleotides of the invention.
X
Q Sequence 1314 BP; 297 A; 369 C; 373 G; 275 T; 0 other;
X
Lignment Scores:
red. No.: 1.47e-120 Length: 1314
core: 1266.00 Matches: 228
ercent Similarity: 98.29% Conservative: 2
est Local Similarity: 97.44% Mismatches: 4
very Match: 79.12% Indels: 0
B: 25 Gaps: 0
S-09-607-745-9 (1-292) x ABX76354 (1-1314)
Y 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
b ::::|||||
610 CGTGTGGTGGTGGGAGGAGGCGCTCTGTGGATCTTGGCCTTGGCAGGTCAGCATCCAG 669
Y 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
b |||||
670 TACGACAAACAGCAGCTGTGTGGAGGAGCATCCTGGACCCCACTGGGTCTCTCAGGCA 729
Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
b |||||
730 GCCCACTGCTTCAGGAACATACCGATGTGTCAACTGGAAGTGGGGCAGGCTCAGAC 789
Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
b |||||
790 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATCATGATTCACACCC 849
Y 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
b |||||
850 ATGTACCCCAAGACAATGACATCGCCCTCATGAGCTGCAGTTCCTCCACTCTCTCTCA 909
Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
b |||||
910 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCCCA 969
Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
b |||||
970 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATATG 1029
Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 210
b |||||
1030 CTGCAGGCGTCAGTCCAGGTTCATTGACAGCACACGGTGCATGATGACAGCATGGTACCAG 1089
Y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThrCys 230

Db 1090 GGGGAAGTCACCGAGAGATGATGTGTGCAGGCATCCCGAAGGGGTGTGGACACCTGC 1149
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
Db 1150 CAGGGTGACAGTGGTGGGCCCTGTATGTACCAATCTGACAGTGGCATGTGGTGGCATC 1209
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
Db 1210 GTTAGCTGGGGCTATGGTGGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1269
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1270 GCCTATCTCAACTGGATCTACAATGTCTTGAAGGCTGAGCTG 1311
RESULT 4
AAH99574/c
ID AAH99574 standard; cDNA; 1854 BP.
XX
AC AAH99574;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human protein encoding cDNA sequence SEQ ID NO:409.
XX
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiant; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovacular; antianaemic; anaemia;
KW antiaggregant; haemostatic; vulnery; antidiabetic; cytostatic;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antitense therapy; vaccine; pancreatitis;
KW antianaphylactic; rheumatoid arthritis; septic shock; inflammation;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder; ss.
XX Homo sapiens.
XX WO200153455-A2.
PN 26-JUL-2001.
XX 22-DEC-2000; 2000WO-US35017.
XX 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Drmanac RT;
PI
XX WPI; 2001-457603/49.
DR P-PSDB; AAM25633.
XX
XX Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
PT
XX
PS Claim 1; Page 506; 1217pp; English.
XX
CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovacular; antianaemic; antiaggregant; haemostatic; vulnery;
CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;

C antiparkinsonian; and immunostimulant. The proteins and polynucleotides
C encoding them can be used in gene therapy, antisense therapy and vaccine
C production. The proteins and polynucleotides are useful for screening for
C agonists or antagonists of a protein and for the treatment and diagnosis
C of disorders associated with the activity of a protein e.g. inflammation,
C rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
C neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
C infections, autoimmunity, genetic diseases, haematopoietic disorders,
C anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
C osteoporosis, severe combined immunodeficiency, eczema, allergic
C rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
C Alzheimer's disease, Parkinson's disease, neurodegenerative and
C neurological disorders.

X Q Sequence 1854 BP; 397 A; 513 C; 495 G; 449 T; 0 other;

Alignment Scores:

red. No.: 2.31e-120 Length: 1854
core: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
B: 22 Gaps: 0

S-09-607-745-9 (1-292) x AAH99574 (1-1854)

Y 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
b 1605 CGTGTGGTGGGAGGAGGCCCTCTGTGGATCTTGGCCTTGGCAGGTCAGCATCCAG 1546
Y 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
b 1545 TACGACAAACAGCACGTCGTGTGGAGGAGCATCTGGACCCCACTGGTCTCTCAGGCA 1486
Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
b 1485 GCCCACTGCTTCAGGAACATACCGATGTGTCACTGGAAGTGGGGCAGGCTCAGAC 1426
Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
b 1425 AAACGGGCAGCTTCCATCCCTGGCTGTGGCCAGATCATCATTAATTAACCC 1366
Y 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
b 1365 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGTCGAGTCCCACTCACTTCTCA 1306
Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
b 1305 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGATGAGGAGCTCACTCCAGCCACCCCA 1246
Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
b 1245 CTCTGGATCATTTGATGGGGCTTTACGAAGCAGACATGGAGGAAGATGTCTGACATCTG 1186
Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
b 1185 CTGCAGCGCTCAGTCCAGGTCTATTGACAGCACACGGTGTCAATGCAGACGATGGTACCAG 1126
Y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThrCys 230
b 1125 GGGGAAGTCACCGAGAAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGACACCTGC 1066
Y 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
b 1065 CAGGGTGACAGTGTGGGCCCTGATGTATCAATCTGACCACTGAGCATGTGGTGGCATC 1006
Y 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
b 1005 GTTAGCTGGGGCTATGGCTCGGGGGCCCGAGCACCCCGAGGAGTATACCAAGGTCTCA 946
Y 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
b 945 GCCTATCTCAACTGGATCTACAATGTCTTGAAGGCTGAGCTG 904

RESULT 5
AAX87154
ID AAX87154 standard; cDNA; 2038 BP.

XX AAX87154;

XX 27-SEP-1999 (first entry)

DE Human protease HUPM-6 cDNA.

KW Serine protease; human; HUPM-6; cell proliferation; cancer;
KW immune disorder; inflammation; therapy; ss.

XX Homo sapiens.

FH Key Location/Qualifiers
FT CDS 200..1507
FT /*tag= a

PN WO9936550-A2.

XX 22-JUL-1999.

XX 12-JAN-1999; 99WO-US00655.

XX 16-JAN-1998; 98US-0008271.

PA (INCY-) INCYTE PHARM INC.

PI Bandman O, Corley NC, Guegler KJ, Hillman JL, Shah P;
PI Tang YT, Yue H;

DR WPI; 1999-430616/36.

DR P-PSDB; AAY06437.

XX Novel human protease molecules useful in the treatment of
XX developmental disorders and/or cancers

PS Claim 8; Page 86-87; 90pp; English.

CC This nucleotide sequence codes for HUPM-6 (see AAY06437), a novel
CC human protease. HUPM-6 cDNA was initially identified in Incyte
CC Clone 1337018 from the colon cDNA library COLNNOT13 using a
CC computer search for amino acid sequence alignments. The present
CC sequence is a consensus sequence derived from overlapping and/or
CC extended nucleic acid sequences: Incyte Clones 1271725 (TESTTUT02),
CC 1337018, 586982 and 588598 (UTRSNOT01). A fragment comprising
CC nucleotides 900-949 of the present sequence can be used for
CC hybridisation. This sequence encompasses an active site residue.
CC Northern analysis shows expression of HUPM-6 in gastrointestinal,
CC and male an female reproductive cDNA libraries. Approximately 65%
CC of these libraries are associated with neoplastic disorders, and
CC 22% with the immune response. The invention provides 12 new human
CC proteases, i.e. HUPM-1 to -12 (see AAY06432-43), and the
CC polynucleotides encoding them (see AAX87149-60). Also provided are
CC vectors, host cells and methods for producing HUPM polypeptides, as
CC well as agonists and antagonists of HUPM. Methods for treating or
CC preventing cell proliferative disorders and immune disorders using
CC HUPM or HUPM antagonists are claimed.

SQ Sequence 2038 BP; 462 A; 591 C; 569 G; 416 T; 0 other;

Alignment Scores:

Pred. No.: 2.61e-120 Length: 2038
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 20 Gaps: 0

US-09-607-745-9 (1-292) x AAX87154 (1-2038)

R 20-OCT-1998; 98US-0104987.
R 20-OCT-1998; 98US-0105000.
R 20-OCT-1998; 98US-0105002.
R 21-OCT-1998; 98US-0105104.
R 22-OCT-1998; 98US-0105169.
R 22-OCT-1998; 98US-0105266.
R 26-OCT-1998; 98US-0105693.
R 26-OCT-1998; 98US-0105694.
R 27-OCT-1998; 98US-0105807.
R 27-OCT-1998; 98US-0105881.
R 27-OCT-1998; 98US-0105882.
R 27-OCT-1998; 98US-0106062.
R 28-OCT-1998; 98US-0106023.
R 28-OCT-1998; 98US-0106029.
R 28-OCT-1998; 98US-0106030.
R 28-OCT-1998; 98US-0106032.
R 28-OCT-1998; 98US-0106033.
R 28-OCT-1998; 98US-0106178.
R 29-OCT-1998; 98US-0106248.
R 29-OCT-1998; 98US-0106384.
R 29-OCT-1998; 98US-0108500.
R 30-OCT-1998; 98US-0106464.
R 03-NOV-1998; 98US-0106856.
R 03-NOV-1998; 98US-0106902.
R 03-NOV-1998; 98US-0106905.
R 03-NOV-1998; 98US-0106919.
R 03-NOV-1998; 98US-0106932.
R 03-NOV-1998; 98US-0106934.
R 10-NOV-1998; 98US-0107783.
R 17-NOV-1998; 98US-0108775.
R 17-NOV-1998; 98US-0108779.
R 17-NOV-1998; 98US-0108787.
R 17-NOV-1998; 98US-0108788.
R 17-NOV-1998; 98US-0108801.
R 17-NOV-1998; 98US-0108802.
R 17-NOV-1998; 98US-0108806.
R 17-NOV-1998; 98US-0108807.
R 17-NOV-1998; 98US-0108867.
R 17-NOV-1998; 98US-0108925.
R 18-NOV-1998; 98US-0108848.
R 18-NOV-1998; 98US-0108849.
R 18-NOV-1998; 98US-0108850.
R 18-NOV-1998; 98US-0108851.
R 18-NOV-1998; 98US-0108852.
R 18-NOV-1998; 98US-0108858.
R 18-NOV-1998; 98US-0108904.
(GETH) GENENTECH INC.

Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
WPI; 2000-237871/20.
P-PSDB; AAY99417.

New mammalian DNA sequences encoding transmembrane, receptor or
secreted PRO polypeptides, useful for screening of potential peptide or
small molecule inhibitors of the relevant receptor/ligand interactions

Claim 2; Fig 155; 773pp; English.

AAA37022 to AAA37144 encode the new isolated human transmembrane,
receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
transmembrane and receptor PRO proteins can be used for screening of
potential peptide or small molecule inhibitors of the relevant
receptor/ligand interactions. The polypeptides and nucleotide sequences
encoding then have various industrial applications, including uses as
pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
PCR primers and hybridisation probes used in the isolation of the PRO
polypeptides from the present invention.

Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

Alignment Scores:

Pred. No.: 2.65e-120 Length: 2063
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 21 Gaps: 0
US-09-607-745-9 (1-292) x AAA37099 (1-2063)
Qy 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
Db 804 CGTGTGGTGGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTCCAGTCCAG 863
Qy 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
Db 864 TACGACAAACAGCAGCTGTGTGGAGGAGCATCTGGACCCCTGGGTCTCTCAGGCA 923
Qy 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
Db 924 GCCCACTGCTTCAGGAACATACCGATGTGTTCACTGGAAGTGGGGCAGGCTCAGAC 983
Qy 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
Db 984 AAACCTGGGAGCTTCCCATCCCTGGCTGGCCCAAGATCATCATGATTGAATTCACCCC 1043
Qy 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
Db 1044 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGTGCAGTCCCACTCACTTTCTCA 1103
Qy 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
Db 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGTCACTCCAGCCACCCCA 1163
Qy 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
Db 1164 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGGAAGATGTCTGACATCTG 1223
Qy 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
Db 1224 CTGCAGGCGTCAGTCCAGGTCAATTGACACACACCGTGCATGACAGATGCGTACCCAG 1283
Qy 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
Db 1284 GGGGAAGTCACCGAAGATGATGTGTGAGGCGATCCCGAAGGGGTGTGGACACCTGC 1343
Qy 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
Db 1344 CAGGGTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGCATC 1403
Qy 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
Db 1404 GTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCCGAGGATATACCAAGGTCTCA 1463
Qy 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1464 GCCTATCTCACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505

RESULT 7

AAS46089

ID AAS46089 standard; cDNA; 2063 BP.

XX AAS46089;

XX 18-DEC-2001 (first entry)

XX Human DNA encoding PRO polypeptide sequence #165.

XX PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW PCR primer.

XX

IS Homo sapiens.
X WO200168848-A2.
N
X 20-SEP-2001.
D
X 28-FEB-2001; 2001WO-US06520.
F
X
X 01-MAR-2000; 2000WO-US05601.
R 02-MAR-2000; 2000WO-US05841.
R 03-MAR-2000; 2000US-187202P.
R 06-MAR-2000; 2000US-186968P.
R 14-MAR-2000; 2000US-189320P.
R 14-MAR-2000; 2000US-189328P.
R 15-MAR-2000; 2000WO-US06884.
R 21-MAR-2000; 2000US-190828P.
R 21-MAR-2000; 2000US-191007P.
R 21-MAR-2000; 2000US-191048P.
R 21-MAR-2000; 2000US-191314P.
R 28-MAR-2000; 2000US-192655P.
R 29-MAR-2000; 2000US-193032P.
R 29-MAR-2000; 2000US-193053P.
R 30-MAR-2000; 2000WO-US08439.
R 04-APR-2000; 2000US-194449P.
R 04-APR-2000; 2000US-194647P.
R 11-APR-2000; 2000US-195975P.
R 11-APR-2000; 2000US-196000P.
R 11-APR-2000; 2000US-196187P.
R 11-APR-2000; 2000US-196690P.
R 11-APR-2000; 2000US-196820P.
R 18-APR-2000; 2000US-198121P.
R 18-APR-2000; 2000US-198585P.
R 25-APR-2000; 2000US-199397P.
R 25-APR-2000; 2000US-199550P.
R 25-APR-2000; 2000US-199654P.
R 03-MAY-2000; 2000US-201516P.
R 17-MAY-2000; 2000WO-US13705.
R 22-MAY-2000; 2000WO-US14042.
R 30-MAY-2000; 2000WO-US14941.
R 02-JUN-2000; 2000WO-US15264.
R 05-JUN-2000; 2000US-209832P.
R 28-JUL-2000; 2000WO-US20710.
R 22-AUG-2000; 2000US-0644848.
R 24-AUG-2000; 2000WO-US23328.
R 08-NOV-2000; 2000WO-US30952.
R 01-DEC-2000; 2000WO-US32678.
R 20-DEC-2000; 2000WO-US34956.
X
X (GETH) GENENTECH INC.
X
X Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
X Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
X WPI; 2001-602746/68.
R P-PSDB; AAU29188.
X
X Novel nucleic acids encoding PRO polypeptides, used to diagnose the
T presence of tumours, such as prostate and breast tumours, in mammals and
T to screen for modulators of the compounds -
X
X Claim 2; Fig 329; 774pp; English.
X
X Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
C primers for PRO polypeptides of the invention. The sequences of the
C invention can be used to detect the presence of a tumour in a mammal by
C comparing the level of expression of a PRO polypeptide in a test sample
C of cells from the animal and a control sample of normal cells, whereby a
C higher level of expression in the test sample indicates the presence of a
C tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
C pigs, goats and rabbits but are preferably human. The polypeptides can be
C used to stimulate tumour necrosis factor (TNF) alpha release from human
C blood, when contacted with it. A specific polypeptide can be used to
C stimulate the proliferation or differentiation of chondrocyte cells. The

CC PRO proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX
SQ Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

Alignment Scores:
Pred. No.: 2.65e-120 Length: 2063
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 22 Gaps: 0

US-09-607-745-9 (1-292) x AAS46089 (1-2063)

QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
Db ::::|||||
804 CGTGTGGTGGTGGGAGGAGGCGCTCTGTGGATTCTTGGCCTGGCAGGTTCAGCATCCAG 863

QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
Db :|||
864 TACGACAAACACACACGCTCTGTGGAGGAGCATCTTGGACCCCTGGTCTCTCAGCGCA 923

QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
Db :|||
924 GCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAAGTGGGCGGAGGCTCAGAC 983

QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
Db :|||
984 AAACCTGGGAGCTTCCATCCCTGGCTGTGGCCAAAGATCATCATCATGAATTCACACCC 1043

QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
Db :|||
1044 ATGTACCCCAAGACAAATGACATCGCCTCATGAAGTGCAGTCCCACTCACTTTCTCA 1103

QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
Db :|||
1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163

QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
Db :|||
1164 CTCTGGATCATGGATGGGCTTACGAAGCAGAAATGGAGGAGAGATGTCTGACATACTG 1223

QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
Db :|||
1224 CTGCAGGCGTCAGTCCAGGTCATTGACAGCACACCGTGCATATGCAGACGATGCGTACCAG 1283

QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
Db :|||
1284 GGGGAAGTCACCGAGAGATGATGTGTGAGGAGCATCCCGAAGGGGGTGTGGACACCTGC 1343

QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
Db :|||
1344 CAGGGTGACAGTGGTGGGCGCCCTGATGTACCAATCTGACAGTGGCATGTGTGGGCGATC 1403

QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
Db :|||
1404 GTTAGCTGGGCTATGGCTGCGGGGGGCGGAGCACCCCGAGGAGTATACACCAAGGTCTCA 1463

QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db :|||
1464 GCCTATCTCACTGGATCTACATGTCTGGAAGGCTGAGCTG 1505

RESULT 8
AAF92113
ID AAF92113 standard; cDNA; 2063 BP.
XX
AC AAF92113;
XX
DT 15-MAY-2001 (first entry)

X	Human PRO1570 cDNA.	Db	924	GCCCACTGCTTCAGGAAACATACCGATGTGTCAACTGGAAGTGCGGGCAGGCTCAGAC	983
E					
X	Human; PRO protein; mapping; ss.	Qy	111	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro	130
W					
X		Db	984	AAACTGGGCAGCTTCCCATCCCTGGCTGGCCCAAGATCATCATCATTAATCAACCCC	1043
X					
S	Homo sapiens.	Qy	131	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	150
X					
N	WO200116318-A2.	Db	1044	ATGTACCCCAAGACAATGACATCGCCCTTCATGAAGCTGCAGTCCCACCTCTCTCTCA	1103
X	08-MAR-2001.				
X		Qy	151	GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	170
F	24-AUG-2000; 2000WO-US23328.	Db	1104	GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA	1163
X					
R	01-SEP-1999; 99WO-US20111.	Qy	171	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	190
R	15-SEP-1999; 99WO-US21090.				
R	07-DEC-1999; 99US-0169495.	Db	1164	CTCTGGATCATGGATGGGGCTTTACGAAGCAGAAATGGAGGAGATGTCTGACATACTG	1223
R	09-DEC-1999; 99US-0170262.				
R	11-JAN-2000; 2000US-0175481.	Qy	191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	210
R	18-FEB-2000; 2000WO-US04341.	Db	1224	CTGCAGCGCTCAGTCCAGGTCAATTGACAGCACACCGGTGCAATGCAGACGATCGGTACCAG	1283
R	18-FEB-2000; 2000WO-US04342.				
R	22-FEB-2000; 2000WO-US04414.	Qy	211	GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	230
R	01-MAR-2000; 2000WO-US05601.	Db	1284	GGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCCGGAAGGGGTGTGGACACCTGC	1343
R	03-MAR-2000; 2000US-0187202.				
R	25-APR-2000; 2000US-0199397.	Qy	231	GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	250
R	22-MAY-2000; 2000WO-US14042.	Db	1344	CAGGGTGACACTGGTGGGCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGGCATC	1403
R	05-JUN-2000; 2000US-0209832.	Qy	251	ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer	270
X	(GETH) GENENTECH INC.	Db	1404	GTTAGCTGGGGCTATGGCTGCGGGGGCCCCGAGCACCCCCAGGAGTATACACCAAGGTCTCA	1463
A	Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;				
X	Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;				
I	WPI; 2001-183260/18.				
I	P-PSDB; AAB87581.				
X					
T	Eighty four nucleic acids encoding PRO polypeptides, useful in				
T	molecular biology, including use as hybridization probes, and in				
T	chromosome and gene mapping. -				
X	Claim 2; Fig 111; 278pp; English.				
S					
X	The present sequence is the coding sequence for a human PRO polypeptide				
C	(secreted and transmembrane). The PRO protein, and PRO agonists, PRO				
C	antagonists or anti-PRO antibodies are useful for preparation of a				
C	medicament useful in the treatment of a condition which is responsive to				
C	the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO				
C	protein may also be employed as molecular weight markers for protein				
C	electrophoresis. The PRO coding sequence has applications in molecular				
C	biology, including use as hybridisation probes, and in chromosome and				
C	gene mapping.				
X					
Q	Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;				
	lignment Scores:				
	red. No.:		2.65e-120	Length:	2063
	core:		1266.00	Matches:	228
	ercent Similarity:		98.29%	Conservative:	2
	est Local Similarity:		97.44%	Mismatches:	4
	uery Match:		79.12%	Indels:	0
	B:		22	Gaps:	0
	S-09-607-745-9 (1-292) x AAF92113 (1-2063)				
Y	51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln	70			
b	804 CGTGTGGTGGTGGGAGGAGGCGCTCTGTGGATTCTTGGCCTTGGCAGGTCAGCATCCAG	863			
Y	71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla	90			
b	864 TACGACAAACAGCACGCTCTGTGAGGGAGCATCTGTGACCCCCCTGGTCTCTCACGCA	923			
Y	91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	110			

RESULT 9
AAF54396

ID AAF54396 standard; DNA; 2063 BP.

AC AAF54396;

DT 02-APR-2001 (first entry)

DE DNA encoding protein of the invention #76.

XX Secreted; transmembrane; gene therapy; ss.

OS Unidentified.

PN WO200078961-A1.

PD 28-DEC-2000.

PF 18-FEB-2000; 2000WO-US04342.

PR 23-JUN-1999; 99US-0141037.

PR 20-JUL-1999; 99US-0144758.

PR 26-JUL-1999; 99US-0145698.

PR 01-SEP-1999; 99WO-US20111.

PR 29-OCT-1999; 99US-0162506.

PR 30-NOV-1999; 99WO-US28313.

PR 02-DEC-1999; 99WO-US28551.

PR 16-DEC-1999; 99WO-US30095.

PR 05-JAN-2000; 2000WO-US00219.

PR 06-JAN-2000; 2000WO-US00376.

(GETH) GENENTECH INC.

Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;

Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

I Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
I Watanabe CK, Williams PM, Wood WI;
X WPI; 2001-071395/08.

X Secreted and transmembrane proteins and nucleic acids designated PRO,
T useful as hybridization probes, in chromosome and gene mapping and gene
T therapy -

X Claim 2; Fig 151; 787pp; English.

X The present invention relates to secreted and transmembrane proteins.
C These proteins and the DNA encoding them may be used as hybridization
C probes, in chromosome and gene mapping and in the generation of
C anti-sense RNA and DNA. They may also be used to generate either
C transgenic animals or knockout animals which are in turn useful for
C development and screening of therapeutically useful reagents.
C The nucleic acids may also be used in gene therapy.

X Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

Alignment Scores:

red. No.: 2.65e-120 Length: 2063
core: 1266.00 Matches: 228
percent Similarity: 98.29% Conservative: 2
est Local Similarity: 97.44% Mismatches: 4
every Match: 79.12% Indels: 0
B: 22 Gaps: 0

S-09-607-745-9 (1-292) x AAF54396 (1-2063)

NY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
b 804 CGTGTGGTGGTGGGAGGAGGCTCTGTGATTCTTGGCCTTGGCAGGTCCAG 863
NY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
b 864 TACGACAAACAGCACGCTCTGTGGAGGAGCATCCTGGACCCCACTGGTCTCACGGCA 923
NY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
b 924 GCCCACTGCTTCAGGAACATACCGATGTGTTCACTGGAAGTGGGGCAGGCTCAGAC 983
NY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
b 984 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATCAATTCAACCC 1043
NY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
b 1044 ATGTACCCCAAGACATGATGATCGCCCTCATGAAGCTGCAGTTCCTCACTTCTCA 1103
NY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
b 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCCCA 1163
NY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu 190
b 1164 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGAAGATGCTGACATCTG 1223
NY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
b 1224 CTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCATGCAGACGATCGTACCAG 1283
NY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThrCys 230
b 1284 GGGGAAGTCACCGAGAGATGATGTGTGCAGGATCCCGAAGGGGTGTGGACACCTGC 1343
NY 231 GlnGlyAspSerGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
b 1344 CAGGGTGACAGTGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGCATC 1403
NY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270

Db 1404 GTTAGCTGGGGCTATGGCTCGGGGGCCCGAGCACCCCGAGGTATACACCAAGGTCTCA 1463
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1464 GCCTATCTCAACTGGATCTACAATGTCTGTGAAGGCTGAGCTG 1505

RESULT 10

ABS74433
ID ABS74433 standard; cDNA; 2063 BP.

XX ABS74433;

XX 10-DEC-2002 (first entry)

DE Human cDNA encoding secreted/transmembrane protein PRO1570.

KW Human; ss; gene; secreted protein; transmembrane protein; antirheumatic;
KW antiarthritic; osteopathic; sports-related joint problem;
KW articular cartilage defect; osteoarthritis; rheumatoid arthritis.

XX Homo sapiens.

XX US2002119130-A1.

XX 29-AUG-2002.

XX 06-DEC-2001; 2001US-0006867.

XX 29-OCT-1997; 97US-063435P.

XX 29-OCT-1997; 97US-064215P.

XX 22-APR-1998; 98US-082797P.

XX 29-APR-1998; 98US-083495P.

XX 15-MAY-1998; 98US-085579P.

XX 10-JUN-1998; 98US-088811P.

XX 10-JUN-1998; 98US-088824P.

XX 10-JUN-1998; 98US-088825P.

XX 11-JUN-1998; 98US-088863P.

XX 12-JUN-1998; 98US-089105P.

XX 16-JUN-1998; 98US-089514P.

XX 16-SEP-1998; 98WO-US19330.

XX 08-MAR-1999; 99WO-US05028.

XX 14-MAY-1999; 99WO-US10733.

XX 02-JUN-1999; 99WO-US12252.

XX 01-SEP-1999; 99WO-US20111.

XX 15-SEP-1999; 99WO-US21090.

XX 15-SEP-1999; 99WO-US21194.

XX 22-DEC-1999; 99WO-US30720.

XX 18-FEB-2000; 2000WO-US04341.

XX 18-FEB-2000; 2000WO-US04342.

XX 30-MAR-2000; 2000WO-US08439.

XX 22-MAY-2000; 2000WO-US14042.

XX 02-JUN-2000; 2000WO-US15264.

XX 23-AUG-2000; 2000WO-US23522.

XX 24-AUG-2000; 2000WO-US23328.

XX 10-NOV-2000; 2000WO-US30873.

XX 01-DEC-2000; 2000WO-US32378.

XX 20-DEC-2000; 2000WO-US34956.

XX 28-FEB-2001; 2001WO-US06520.

XX 20-JUN-2001; 2001WO-US19692.

XX 29-JUN-2001; 2001WO-US21066.

XX 09-JUL-2001; 2001WO-US21735.

(GETH) GENENTECH INC.

XX Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;

XX Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;

XX WPI; 2002-731348/79.

XX P-PSDB; ABG95906.

XX New isolated secreted and transmembrane PRO polypeptide useful for
PT modulating biological activity of a cell, or for treating
PT sports-related joint problems, osteoarthritis or rheumatoid arthritis

X S Claim 2; Fig 111; 399pp; English.

X C The invention relates to an isolated secreted and transmembrane PRO

X C polypeptide having 80 % sequence identity to a sequence appearing

X C as ABG95851-ABG95934 or their associated signal peptide, or a sequence of

X C an extracellular domain of the proteins with their associated signal

X C peptide or lacking its associated signal peptide. Also included are

X C the nucleic acids encoding the proteins, vectors, host cells,

X C fusion proteins and antibodies which specifically bind to the proteins.

X C The proteins are useful for detecting a polypeptide designated as A, B, C

X C or D in a sample suspected of containing an A, B, C or D polypeptide,

X C by contacting the sample with a polypeptide designated as E, F, G, H or

X C I (or vice versa) and determining the formation of a A/E, B/F, B/G, C/H

X C or D/I polypeptide conjugate in the sample, where the formation of the

X C conjugate is indicative of the presence of an A, B, C or D polypeptide

X C in the sample, where A is a PRO10272 polypeptide, B is a PRO20110

X C polypeptide, C is a PRO10096 polypeptide, D is a PRO19760 polypeptide,

X C E is a PRO5801 polypeptide, F is a PRO1 polypeptide, G is a PRO20040

X C polypeptide, H is a PRO20233 polypeptide and I is a PRO1890

X C polypeptide. The sample comprises a cell suspected of expressing the A,

X C B, C or D polypeptide. The E, F, G, H or I polypeptide is labeled with

X C a detectable label or is attached to a solid support. The proteins are

X C useful for linking a bioactive molecule to a cell expressing a

X C polypeptide designated as A, B, C or D or E, F, G, H or I. The bioactive

X C molecule is a toxin, a radiolabel or an antibody. The bioactive molecule

X C causes death of the cell. A, B, C, D, E, F, G, H, or I, or antibodies

X C against them are useful for modulating a biological activity of a cell

X C expressing a polypeptide designated as A, B, C or D or E, F, G, H, or

X C I. The cell is killed. The proteins are useful for identifying

X C agonists or antagonists, for the preparation of a medicament useful in

X C the treatment of a condition which is responsive to the proteins, as

X C molecular weight markers for protein electrophoresis purposes, and as

X C therapeutic agents for treating sports-related joint problems,

X C articular cartilage defects, osteoarthritis or rheumatoid arthritis.

X C Nucleic acids encoding the proteins are useful as hybridisation probes,

X C in chromosome and gene mapping, in the generation of anti-sense RNA and

X C DNA, for the preparation of the proteins, to generate transgenic or

X C knockout animals which are useful in the development and screening of

X C therapeutic useful reagents, for chromosome identification, and in gene

X C therapy. The antibody is useful as a therapeutic agent, in a diagnostic

X C assay and for affinity purification of the protein from recombinant

X C cell culture natural sources. The present sequence encodes a novel

X C secreted or transmembrane protein of the invention.

X Q Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

Alignment Scores:

red. No.:	2.65e-120	Length:	2063
core:	1266.00	Matches:	228
ercent Similarity:	98.29%	Conservative:	2
est Local Similarity:	97.44%	Mismatches:	4
very Match:	79.12%	Indels:	0
B:	24	Gaps:	0

IS-09-607-745-9 (1-292) x ABS74433 (1-2063)

Y 51 LysIleValGlyGlyTyxAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70

b 804 CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGGTGAGTCCAG 863

Y 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90

b 864 TACGACAAACAGACACGTCGTGTGGAGGAGCATCTCTGGACCCCTGCTCAGGGCA 923

Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110

b 924 GCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGCTCAGAC 983

Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130

b 984 AAACGGGAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATCTGATTCAACCCC 1043

QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150

Db 1044 ATGTACCCCAAGACATGACATCGCCCTCATGAAGTCGAGTCCCACTCTCTCTCA 1103

QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170

Db 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGTCACTCAGCCACCCCA 1163

QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190

Db 1164 CTCTGGATCATTTGGTGGGCTTTACGAAGCAGATGGAGGAAGATGCTGTGACATCTG 1223

QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 210

Db 1224 CTGCAGGCGTCAGTCCAGTCAATTGACAGCACACGGTGCATGCAGACCATGCGTACCAG 1283

QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230

Db 1284 GGGGAAGTCACCGAGAGATGATGTGTGCAGGCATCCCGGAAGGGGTGTGGACACCTGC 1343

QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250

Db 1344 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAATCTGCGTGGCATGTGGTGGCATC 1403

QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270

Db 1404 GTTAGCTGGGGCTATGGCTGCGGGGCGCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1463

QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284

Db 1464 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGTGAGCTG 1505

RESULT 11

ABK11090

ID ABK11090 standard; cDNA; 2063 BP.

XX AC ABK11090;

XX DT 05-JUN-2002 (first entry)

XX DE cDNA encoding tumour-associated antigenic target protein, TAT135.

XX KW TAT135; Tumour-associated Antigenic Target; tumour;

XX KW breast cancer; colorectal cancer; lung cancer; ovarian cancer;

XX KW central nervous system cancer; liver cancer; bladder cancer; melanoma;

XX KW pancreatic cancer; leukaemia; gene therapy; gene; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 210..1508

XX FT /*tag= a

XX FT /product= "TAT135"

XX FT /note= "Tumour-associated antigenic target"

XX PN WO200216429-A2.

XX PD 28-FEB-2002.

XX PF 22-JUN-2001; 2001WO-US20118.

XX PR 24-AUG-2000; 2000WO-US23328.

XX PR 26-SEP-2000; 2000US-235451P.

XX PR 01-DEC-2000; 2000WO-US32678.

XX PR 28-FEB-2001; 2001WO-US06520.

XX PR 01-MAR-2001; 2001WO-US06666.

XX PA (GETH) GENENTECH INC.

XX PI Goddard A, Godowski PJ, Gurney AL, Hillan KJ, Polakis P, Smith V;

XX PI Wood WI, Wu TD, Zhang Z;

XX DR WPI; 2002-280917/32.

R P-PSDB; AAU76535.
X Novel isolated tumour-associated antigenic target polypeptides which are
T useful as targets for cancer therapy and diagnosis in mammals -
X
S Claim 1; Fig 2; 121pp; English.
X
C The invention relates to an isolated tumour-associated antigenic target
C polypeptide (TAT) (I), specifically TAT134-TAT138 polypeptides, and the
C polynucleotides (II) encoding them. (II) is useful for diagnosing the
C presence of a tumour in a mammal, where the level of expression of (II)
C is indicative on the presence of tumour in the mammal from which the
C test sample was obtained. Antibody to (I) is useful for killing a cancer
C cell (e.g. breast cancer cell, a colorectal cancer cell, a lung cancer
C cell, an ovarian cancer cell, a central nervous system (CNS) cancer cell,
C a liver cancer cell, a bladder cancer cell, a pancreatic cancer cell, a
C melanoma cell or a leukaemia cell) that expresses (I). Oligonucleotides
C hybridising to (II) are useful as diagnostic probes, antisense
C oligonucleotide probes or for encoding fragments of full length TAT
C polypeptide. (II) is also useful in chromosome and gene mapping and in
C the generation of antisense RNA and DNA probes, for constructing
C hybridisation probes for mapping the gene encoding TAT and for genetic
C analysis of individuals with genetic disorders. (II) is also useful for
C generating either transgenic animals or knockout animals, and in gene
C therapy. The TAT polypeptides and nucleic acids may also be used for
C tissue typing and the TAT polypeptides are useful for screening
C compounds that mimic the TAT polypeptide (agonist) or prevent the
C effect of TAT polypeptide (antagonist). The antibody is useful for
C staging TAT polypeptide-expressing cancers, purifying or
C immunoprecipitating TAT polypeptide from cells, for detection and
C quantitation of TAT polypeptide in vitro, e.g., in an enzyme linked
C immunosorbent assay (ELISA) or Western blot. The antibodies are also
C useful for treating a TAT-expressing cancer or alleviating one or more
C symptoms of cancer in a mammal. The present sequence represents the
C coding sequence of TAT135.
X
Q Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

Alignment Scores:
red. No.: 2.65e-120 Length: 2063
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
JB: 24 Gaps: 0

US-09-607-745-9 (1-292) x ABK11090 (1-2063)
Y 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
b 804 CGTGTGGTGGTGGGGAGGAGGCTCTGTGGATCTTGGCCTTGGCAGGTCCAGTCCAG 863
Y 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
b 864 TACGACAAACAGCACGTCGTGTGGAGGAGCATCCTGGACCCCACTGGTCTCAGCGCA 923
Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
b 924 GCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGTGGGGCAGGCTCAGAC 983
Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
b 984 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCAAGATCATCATCATGATTCACACCC 1043
Y 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
b 1044 ATGTACCCCAAGACATGACATCGCCCTCATGAAGCTGCAGTCCCACTCACTTCTCA 1103
Y 151 GlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThrPro 170
b 1104 GGCACAGTCAGGCCCACTGTGTCTGCCCTTCTTTGATGAGGAGTCACTCCAGCCACCCCA 1163
Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu 190

Db 1164 CTCTGGATCATGGATGGGCTTTACGAAGCAGAATGGAGGAAGATGTCTGACATCTG 1223
Qy 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 210
Db 1224 CTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGCATGTGAGACGATCGGTACCAG 1283
Qy 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
Db 1284 GGGGAAGTCCCGAGAAGATGATGTGTGAGGATCCCGGAGGGGTGTGGACACCTGC 1343
Qy 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
Db 1344 CAGGGTGACAGTGGTGGCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGGCATC 1403
Qy 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
Db 1404 GTTAGCTGGGGCTATGGCTGCGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1463
Qy 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1464 GCCTATCTCAACTGGATCTACAATGTGTGGAAGGTGAGCTG 1505
RESULT 12
ACA57847
ID ACA57847 standard; cDNA; 2063 BP.
XX
AC ACA57847;
XX
DT 10-JUN-2003 (first entry)
XX
DE Human PRO1570 cDNA.
XX
KW Human; PRO; secreted; transmembrane; cytostatic; TNF-alpha; blood; gene;
KW tumour necrosis factor alpha release; chondrocyte cell; proliferation;
KW differentiation; tumour; gene therapy; ss.
OS Homo sapiens.
PN US2003036143-A1.
XX
PD 20-FEB-2003.
XX
PF 02-JUL-2002; 2002US-0187600.
XX
PR 16-SEP-1998; 98WO-US19330.
PR 07-OCT-1998; 98WO-US21141.
PR 01-DEC-1998; 98WO-US25108.
PR 08-MAR-1999; 99WO-US05028.
PR 14-MAY-1999; 99WO-US10733.
PR 02-JUN-1999; 99WO-US12252.
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 01-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28551.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 15-MAR-2000; 2000WO-US06884.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.

R	20-DEC-2000;	2000WO-US34956.	PR	04-JUN-1998;	98US-088326P.
R	28-FEB-2001;	2001WO-US06520.	PR	05-JUN-1998;	98US-088167P.
R	01-JUN-2001;	2001WO-US17800.	PR	05-JUN-1998;	98US-088202P.
R	20-JUN-2001;	2001WO-US19692.	PR	05-JUN-1998;	98US-088212P.
R	29-JUN-2001;	2001WO-US21066.	PR	05-JUN-1998;	98US-088217P.
R	09-JUL-2001;	2001WO-US21735.	PR	09-JUN-1998;	98US-088655P.
R	29-AUG-2001;	2001WO-US27099.	PR	10-JUN-1998;	98US-088722P.
R	18-SEP-1997;	97US-059263P.	PR	10-JUN-1998;	98US-088738P.
R	18-SEP-1997;	97US-059266P.	PR	10-JUN-1998;	98US-088740P.
R	17-OCT-1997;	97US-062250P.	PR	10-JUN-1998;	98US-088811P.
R	21-OCT-1997;	97US-063486P.	PR	10-JUN-1998;	98US-088824P.
R	24-OCT-1997;	97US-063120P.	PR	10-JUN-1998;	98US-088825P.
R	24-OCT-1997;	97US-063121P.	PR	10-JUN-1998;	98US-088826P.
R	28-OCT-1997;	97US-063540P.	PR	11-JUN-1998;	98US-088861P.
R	28-OCT-1997;	97US-063541P.	PR	11-JUN-1998;	98US-088863P.
R	28-OCT-1997;	97US-063544P.	PR	11-JUN-1998;	98US-088876P.
R	28-OCT-1997;	97US-063564P.	PR	12-JUN-1998;	98US-089090P.
R	29-OCT-1997;	97US-063734P.	PR	12-JUN-1998;	98US-089105P.
R	31-OCT-1997;	97US-063870P.	PR	16-JUN-1998;	98US-089512P.
R	31-OCT-1997;	97US-064103P.	PR	16-JUN-1998;	98US-089514P.
R	13-NOV-1997;	97US-065311P.	PR	17-JUN-1998;	98US-089538P.
R	21-NOV-1997;	97US-066120P.	PR	17-JUN-1998;	98US-089598P.
R	24-NOV-1997;	97US-066466P.	PR	17-JUN-1998;	98US-089653P.
R	24-NOV-1997;	97US-066772P.	PR	18-JUN-1998;	98US-089908P.
R	11-DEC-1997;	97US-069335P.	PR	19-JUN-1998;	98US-089952P.
R	12-DEC-1997;	97US-069425P.	PR	22-JUN-1998;	98US-090246P.
R	17-DEC-1997;	97US-069870P.	PR	22-JUN-1998;	98US-090252P.
R	18-DEC-1997;	97US-068017P.	PR	22-JUN-1998;	98US-090254P.
R	10-MAR-1998;	98US-077450P.	PR	24-JUN-1998;	98US-090429P.
R	11-MAR-1998;	98US-077632P.	PR	24-JUN-1998;	98US-090435P.
R	11-MAR-1998;	98US-077649P.	PR	24-JUN-1998;	98US-090444P.
R	20-MAR-1998;	98US-078866P.	PR	24-JUN-1998;	98US-090461P.
R	20-MAR-1998;	98US-078939P.	PR	24-JUN-1998;	98US-090535P.
R	27-MAR-1998;	98US-079664P.	PR	25-JUN-1998;	98US-090540P.
R	27-MAR-1998;	98US-079786P.	PR	25-JUN-1998;	98US-090676P.
R	31-MAR-1998;	98US-080107P.	PR	25-JUN-1998;	98US-090678P.
R	31-MAR-1998;	98US-080194P.	PR	25-JUN-1998;	98US-090688P.
R	01-APR-1998;	98US-080327P.	PR	25-JUN-1998;	98US-090690P.
R	01-APR-1998;	98US-080333P.	PR	25-JUN-1998;	98US-090694P.
R	08-APR-1998;	98US-081049P.	PR	25-JUN-1998;	98US-090695P.
R	08-APR-1998;	98US-081070P.	PR	25-JUN-1998;	98US-090696P.
R	09-APR-1998;	98US-081195P.	PR	26-JUN-1998;	98US-090862P.
R	15-APR-1998;	98US-081838P.	PR	26-JUN-1998;	98US-090863P.
R	21-APR-1998;	98US-082568P.	PR	26-JUN-1998;	98US-091010P.
R	21-APR-1998;	98US-082569P.	PR	01-JUL-1998;	98US-091359P.
R	22-APR-1998;	98US-082704P.	PR	01-JUL-1998;	98US-091544P.
R	22-APR-1998;	98US-082797P.	PR	02-JUL-1998;	98US-091478P.
R	28-APR-1998;	98US-083322P.	PR	02-JUL-1998;	98US-091486P.
R	29-APR-1998;	98US-083495P.	PR	02-JUL-1998;	98US-091626P.
R	29-APR-1998;	98US-083496P.	PR	02-JUL-1998;	98US-091628P.
R	29-APR-1998;	98US-083499P.	PR	02-JUL-1998;	98US-091632P.
R	29-APR-1998;	98US-083559P.	PR	24-JUL-1998;	98US-094006P.
R	05-MAY-1998;	98US-084366P.	PR	04-AUG-1998;	98US-095282P.
R	06-MAY-1998;	98US-084414P.	PR	10-AUG-1998;	98US-095998P.
R	07-MAY-1998;	98US-084639P.	PR	10-AUG-1998;	98US-096012P.
R	07-MAY-1998;	98US-084640P.	PR	17-AUG-1998;	98US-096757P.
R	07-MAY-1998;	98US-084643P.	PR	17-AUG-1998;	98US-096766P.
R	15-MAY-1998;	98US-085579P.	PR	17-AUG-1998;	98US-096867P.
R	15-MAY-1998;	98US-085580P.	PR	17-AUG-1998;	98US-096891P.
R	15-MAY-1998;	98US-085582P.	PR	17-AUG-1998;	98US-096897P.
R	15-MAY-1998;	98US-085700P.	PR	18-AUG-1998;	98US-096949P.
R	18-MAY-1998;	98US-086023P.	PR	18-AUG-1998;	98US-096959P.
R	22-MAY-1998;	98US-086392P.	PR	18-AUG-1998;	98US-097022P.
R	22-MAY-1998;	98US-086486P.	PR	26-AUG-1998;	98US-097952P.
R	28-MAY-1998;	98US-087098P.	PR	26-AUG-1998;	98US-097954P.
R	28-MAY-1998;	98US-087208P.	PR	26-AUG-1998;	98US-097955P.
R	02-JUN-1998;	98US-087609P.	PR	26-AUG-1998;	98US-097971P.
R	02-JUN-1998;	98US-087759P.	PR	26-AUG-1998;	98US-098014P.
R	03-JUN-1998;	98US-087827P.	PR	01-SEP-1998;	98US-098716P.
R	04-JUN-1998;	98US-088025P.	PR	01-SEP-1998;	98US-098723P.
R	04-JUN-1998;	98US-088028P.	PR	02-SEP-1998;	98US-098803P.
R	04-JUN-1998;	98US-088029P.	PR	02-SEP-1998;	98US-098821P.

```
R 02-SEP-1998; 98US-098843P.
R 09-SEP-1998; 98US-099602P.
R 10-SEP-1998; 98US-099741P.
R 10-SEP-1998; 98US-099754P.
R 10-SEP-1998; 98US-099763P.
R 10-SEP-1998; 98US-099812P.

Alignment Scores:
cored. No.: 2.65e-120 Length: 2063
core: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
3: 25 Gaps: 0

5-09-607-745-9 (1-292) x ACA57847 (1-2063)
Y 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
: : : : :
O 804 CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTGGCAGGTCAGCATCCAG 863
Y 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
: : : : :
O 864 TACGACAAACAGCAGCAGTCTGTGGAGGAGCATCTGGACCCCACTGGTCTCTCAGGCA 923
Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
: : : : :
O 924 GCCCACTGCTTCAGGAACATACCGATGTGTTCACTGGAAGGTGCGGCGAGGTCAGAC 983
Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
: : : : :
O 984 AAACGGGAGCTTCCCATCCCTGCTGGCCCAAGATCATCATATTGAATTCACCC 1043
Y 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
: : : : :
O 1044 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGTGCAGTCCCACCTCACTTCTCA 1103
Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
: : : : :
O 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGTCACTCCAGCCACCCCA 1163
Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
: : : : :
O 1164 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATGGAAGGAGATGTCGTACATACCTG 1223
Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
: : : : :
O 1224 CTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGCTGCAATGCAGACGATCGGTACCAG 1283
Y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
: : : : :
O 1284 GGGGAAGTCACCGAGAGATGATGTGTGCAGGCATCCCGGAGGGGTGTGGACACCTGC 1343
Y 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
: : : : :
O 1344 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACAGTGGCATGTGGTGGGCATC 1403
Y 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
: : : : :
O 1404 GTTAGCTGGGCTATGGTGTGGGGGCCGAGCACCCAGAGTATACACCAAGGTCTCA 1463
Y 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
: : : : :
O 1464 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505
```

```

RESULT 13
D ACA58865 standard; cDNA; 2063 BP.
X ACA58865;
X ACA58865;
X 10-JUN-2003 (first entry)
X
```

```
DE CDNA encoding human secreted polypeptide PRO1570.
XX Human; ss; gene; gene therapy; tumour; cancer.
XX Homo sapiens.
XX US2003013855-A1.
XX 16-JAN-2003.
XX 03-MAY-2002; 2002US-0063616.
XX 30-DEC-1998; 98KR-0062142.
XX 08-MAR-1999; 99WO-US05028.
XX 14-MAY-1999; 99WO-US10733.
XX 30-DEC-1999; 99WO-US31274.
XX 18-FEB-2000; 2000WO-US04341.
XX 01-MAR-2000; 2000WO-US05601.
XX 02-MAR-2000; 2000WO-US05841.
XX 21-MAR-2000; 2000WO-US07532.
XX 22-MAY-2000; 2000WO-US14042.
XX 02-JUN-2000; 2000WO-US15264.
XX 24-AUG-2000; 2000WO-US23328.
XX 10-NOV-2000; 2000WO-US30873.
XX 01-DEC-2000; 2000WO-US32678.
XX 20-DEC-2000; 2000WO-US34956.
XX 28-FEB-2001; 2001WO-US06520.
XX 01-JUN-2001; 2001WO-US17800.
XX 14-MAY-1999; 99US-0311832.
XX 25-AUG-1999; 99US-0380137.
XX 25-AUG-1999; 99US-0380138.
XX 25-AUG-1999; 99US-0380139.
XX 25-AUG-1999; 99US-0380142.
XX 15-SEP-1999; 99US-0397342.
XX 18-OCT-1999; 99US-0403297.
XX 12-NOV-1999; 99US-0423844.
XX 22-AUG-2000; 2000US-0644848.
XX 18-SEP-2000; 2000US-0664610.
XX 18-SEP-2000; 2000US-0665350.
XX 08-NOV-2000; 2000US-0709238.
XX 20-DEC-2000; 2000US-0747259.
XX 22-MAR-2001; 2001US-0816744.
XX 10-MAY-2001; 2001US-0854208.
XX 10-MAY-2001; 2001US-0854280.
XX 30-MAY-2001; 2001US-0870574.
XX 05-JUN-2001; 2001US-0874503.
XX 29-JUN-2001; 2001US-0869599.
XX 18-JUL-2001; 2001US-0908827.
XX 06-DEC-2001; 2001US-0006867.
XX (GETH ) GENENTECH INC.
XX Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
XX Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;
XX WPI; 2003-330485/31.
XX P-PSDB; ABU71561.
XX New isolated antibody specifically binding a PRO polypeptide, useful
XX for the preparation of a medicament for treating disorders with the
XX aberrant expression or activity of the PRO polypeptide, such as tumor
XX conditions and cancer -
XX Example 4; Page 184-185; 406pp; English.
XX The invention relates to an antibody that binds to a polypeptide with a
XX fully defined sequence given in the specification. The methods and
XX compositions (containing antibodies that specifically bind a PRO
XX polypeptide) of the present invention are useful for the preparation of a
XX medicament for the treatment of disorders associated with the aberrant
XX expression or activity of the PRO polypeptide, such as tumour conditions
XX and cancer. They can also be used to generate transgenic or knockout
XX animals useful in the development and screening of therapeutically useful
```

reagents. The PRO polypeptides and encoding nucleic acids can be used as molecular weight markers for protein electrophoresis, chromosome identification and tissue typing. The PRO polypeptides are useful to induce angiogenesis e.g wound healing; in the treatment of sports-related joint problems, articular cartilage defects, osteoarthritis or rheumatoid arthritis; diabetes; hyperinsulinaemia and hypoinsulinaemia. The antibodies may be used in various diagnostic, competitive binding and/or immunoprecipitation assays. The present sequence represents a cDNA encoding a PRO polypeptide of the invention.

Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

Alignment Scores:		Length:		2063	
ed. No.:	2.65e-120	Matches:	228		
Percent Similarity:	1266.00	Conservative:	2		
Best Local Similarity:	98.29%	Mismatches:	4		
Query Match:	97.44%	Indels:	0		
3:	25	Gaps:	0		
5-09-607-745-9 (1-292) x ACA58865 (1-2063)					
51	LysileValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln	70			
804	CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCCTGGCAGGTCCAGTCCAG	863			
71	TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla	90			
864	TACGACAAACAGCACGCTCTGTGAGGGAGCATCTGTGACCCCACTGGGTCTCAGGCA	923			
91	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	110			
924	GCCCACTGCTTCAGGAACATACCGATGTGTTCACTGGAAGGTGGGGCAGGCTCAGAC	983			
111	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro	130			
984	AAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATATTGAATCAACCCC	1043			
131	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	150			
1044	ATGTACCCCAAGACAATGATGATGCGCCCTCATGAAGCTGCAGTTCCTCACTTCTCA	1103			
151	GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	170			
1104	GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCACCCCA	1163			
171	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	190			
1164	CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGAGGGAAGATGCTGACATACTG	1223			
191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	210			
1224	CTGCAGGCGTCAGTCCAGGTTCATTGACAGCACACGGTGCAATGCAGACGATGGTACCAG	1283			
211	GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	230			
1284	GGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGAAGGGGTGTGGACACCTGC	1343			
231	GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	250			
1344	CAGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAAGTGGCATGTGTGGGCATC	1403			
251	ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer	270			
1404	GTTAGTGGGGCTATGGTGGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTCA	1463			
271	AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	284			
1464	GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG	1505			

SULT 14
ACA60418 standard; cDNA; 2063 BP.

XX	ACA60418;				
AC	11-JUN-2003 (first entry)				
XX	Novel human secreted and transmembrane protein PRO1570 cDNA.				
DT	Human; secreted and transmembrane polypeptide; gene; ss.				
XX	chromosome mapping; gene mapping; transgenic animal; knockout animal;				
DE	therapeutic agent screening; chromosome identification; tissue typing;				
XX	gene therapy.				
KW	Homo sapiens.				
KW	US2003018183-A1.				
KW	23-JAN-2003.				
XX	01-MAY-2002; 2002US-0063512.				
XX	06-DEC-2001; 2001US-0006867.				
XX	(GETH) GENENTECH INC.				
PA	Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;				
XX	Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;				
PI	WPI; 2003-330984/31.				
PI	P-PSDB; ABU72007.				
DR	New secreted and transmembrane PRO polypeptides and nucleic acid				
DR	molecules encoding the polypeptides, useful in gene therapy or				
XX	preparing a medicament for treating a condition that is responsive to				
XX	the PRO polypeptide or antibody				
PS	Disclosure; Fig 111; 409pp; English.				
XX	The invention describes novel isolated PRO polypeptides. The PRO				
CC	polypeptides or anti-PRO antibodies are useful in preparing a medicament				
CC	for treating a condition that is responsive to the PRO polypeptide or				
CC	antibody. The PRO nucleotide sequences may be used as hybridisation				
CC	probes in chromosome and gene mapping, or in generating antisense RNA				
CC	and DNA. PRO nucleic acids are also useful in preparing PRO polypeptides,				
CC	in assays to identify other proteins or molecules involved in binding				
CC	reaction, to generate transgenic animals or knockout animals, which in				
CC	turn are useful in the development and screening of therapeutically				
CC	useful reagents, for chromosome identification, and tissue typing. The				
CC	PRO polypeptides and nucleic acid molecules are also useful in gene				
CC	therapy, and as molecular weight markers for protein electrophoresis				
CC	purposes. The anti-PRO antibodies may be used in diagnostic assays for				
CC	PRO, or for the affinity purification of PRO from recombinant cell				
CC	culture or natural sources. This sequence encodes a novel human				
CC	secreted and transmembrane PRO polypeptide.				
XX	Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;				
Alignment Scores:					
Pred. No.:	2.65e-120	Length:	2063		
Score:	1266.00	Matches:	228		
Percent Similarity:	98.29%	Conservative:	2		
Best Local Similarity:	97.44%	Mismatches:	4		
Query Match:	97.12%	Indels:	0		
DB:	25	Gaps:	0		
US-09-607-745-9 (1-292) x ACA60418 (1-2063)					
Qy	51	LysileValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln	70		
Db	804	CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCCTGGCAGGTCCAGTCCAG	863		
Qy	71	TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla	90		
Db	864	TACGACAAACAGCACGCTCTGTGAGGGAGCATCTGTGACCCCACTGGGTCTCAGGCA	923		

91 AlaHisCysPheArgGlySHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
 924 GCCCACTGCTTCAGGAACATACCGATGTGTCAACTGGAAGTGGGGCAGGCTCAGAC 983
 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
 984 AAAC TGGCAGCTTCCCATCCCTGGCTGTGGCCAAGATCATCATCATGAATTCAACCC 1043
 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
 1044 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCACACTCATTTCCTCA 1103
 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163
 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
 1164 CTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAATGGAGGAAGATGTCTGACATACTG 1223
 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
 1224 CTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCAATGCAGACGATGCGTACCAG 1283
 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
 1284 GGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGAAGGGGGTGTGGACACCTGC 1343
 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
 1344 CAGGGTGACAGTGTGGGGCCCTGTATGTACCAATCTGACCAGTGCATGTGGTGGGCATC 1403
 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
 1404 GTTAGCTGGGGCTATGGCTGCGGGGGCCGAGCACCCCGAGGAGTATACACCAAGGTCTCA 1463
 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
 1464 GCCTATCTCACTGATCTACAAATGCTTGGAAAGGCTGAGCTG 1505

polypeptide, useful for detecting and purifying the polypeptide and also for treating conditions responsive to the antibody -

Disclosure; Fig 111; 408pp; English.

The present invention relates to the isolation of novel human PRO polypeptides, and the polynucleotide sequences encoding them. The PRO polypeptides are secreted and transmembrane proteins. The PRO polypeptides and polynucleotides are useful for preparing a medicament useful in the treatment of a condition responsive to anti-PRO antibody. Anti-PRO antibodies are useful in diagnostic assays for PRO, by detecting its expression in specific cells, tissues or serum, and for affinity purification of PRO from recombinant cell culture or natural sources. ACA63373-ACA63456 represent cDNA sequences encoding the human PRO polypeptides of the invention.

SQ Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

Alignment Scores:		
Pred. No.:	2.65e-120	Length: 2063
Score:	1266.00	Matches: 228
Percent Similarity:	98.29%	Conservative: 2
Best Local Similarity:	97.44%	Mismatches: 4
Query Match:	79.12%	Indels: 0
DB:	25	Gaps: 0

US-09-607-745-9 (1-292) x ACA63428 (1-2063)

QY	51	Lys	Ile	Val	Gly	Gly	Tyr	Ala	Leu	Asp	Val	Asp	Ser	Trp	Pro	Trp	Gln	Val	Ser	Ile	Gln	70	
DB	804	CGT	GTG	TGG	TGG	GAG	GAG	GGC	CTC	TGT	GGA	TTC	TGT	GCC	CTT	TGG	CAG	GTC	CAG	CA	TCC	AG	863
QY	71	Tyr	Asp	Lys	Gln	His	Val	Cys	Gly	Gly	Ser	Ile	Leu	Asp	Pro	His	Trp	Val	Leu	Thr	Ala	90	
DB	864	TAC	GAC	AAA	CAG	CAC	AGT	CTG	TGG	AGG	GAG	CAT	CCT	TGG	ACCC	CCC	CAC	TGG	GT	CCT	CAC	GGCA	923
QY	91	Ala	His	Cys	Phe	Arg	Lys	His	Thr	Asp	Val	Phe	Asn	Trp	Lys	Val	Arg	Ala	Gly	Ser	Asp	110	
DB	924	GCCC	ACT	GTCT	CAG	GAAC	ATAC	CCG	ATG	TTC	AACT	GGA	AGGT	GCG	GGC	AGG	CTC	AG	C			983	
QY	111	Lys	Leu	Gly	Ser	Phe	Pro	Ser	Leu	Ala	Val	Ala	Lys	Ile	Ile	Ile	Ile	Glu	Phe	Asn	Pro	130	
DB	984	AACT	TGG	GAG	CTT	CCC	ATCCC	TGG	CTC	CTG	GCC	AAAG	ATCAT	CAT	CAT	CAT	CAT	TGA	ATT	CA	CCCC	1043	
QY	131	Met	Tyr	Pro	Lys	Asp	Asn	Asp	Ile	Ala	Leu	Met	Lys	Leu	Gln	Phe	Pro	Leu	Thr	Phe	Ser	150	
DB	1044	ATG	TAC	CCC	CAA	AGA	CAAT	GAC	ATCG	CCCT	CAT	GAA	GCTG	CAG	TTC	CC	ACT	CAC	TTT	CT	CA	1103	
QY	151	Gly	Thr	Val	Arg	Pro	Ile	Cys	Leu	Pro	Phe	Phe	Asp	Glu	Glu	Leu	Thr	Pro	Ala	Thr	Pro	170	
DB	1104	GGC	ACAG	TCA	GG	CCC	ATCT	GT	CTG	CCCT	TCT	TGAT	GAG	GAG	CTC	ACT	CC	AG	CC	AC	CCA	1163	
QY	171	Leu	Trp	Ile	Ile	Gly	Trp	Gly	Phe	Thr	Lys	Gln	Asn	Gly	Gly	Lys	Met	Ser	Asp	Ile	Leu	190	
DB	1164	CTC	TGG	ATC	ATT	GAT	TGG	GCTTT	ACG	AAG	CAGA	ATGG	AGG	GAAG	ATGT	CT	G	CA	TACT	AT	CTG	1223	
QY	191	Leu	Gln	Ala	Ser	Val	Gln	Val	Ile	Asp	Ser	Thr	Arg	Cys	Asn	Ala	Asp	Asp	Ala	Tyr	Gln	210	
DB	1224	CTG	CAG	GCG	TCA	GTCC	AGGT	CA	TTG	AC	AG	CAC	AC	GGT	GCA	ATG	C	AG	CG	ATG	CGT	1283	
QY	211	Gly	Glu	Val	Thr	Glu	Lys	Met	Met	Cys	Ala	Gly	Ile	Pro	Glu	Gly	Gly	Val	Asp	Thr	Cys	230	
DB	1284	GGG	GAAG	TCA	CCG	AGA	AGAT	GTG	TG	CAG	GCAT	CC	CGA	AGG	GGGT	GTG	GA	CAC	CT	GC		1343	
QY	231	Gln	Gly	Asp	Ser	Gly	Pro	Leu	Met	Tyr	Gln	Ser	Asp	Gln	Trp	His	Val	Val	Gly	Ile		250	
DB	1344	CAG	GTG	AC	AGT	GTG	TGG	GGCC	CTG	ATG	TAC	CAAT	CT	GAC	CA	AGT	GG	CA	TGT	GGT	GGG	1403	
QY	251	Val	Ser	Trp	Gly	Tyr	Gly	Cys	Gly	Gly	Pro	Ser	Thr	Pro	Gly	Val	Tyr	Thr	Lys	Val	Ser	270	
DB	1404	GTT	AGCT	TGG	GGCT	AT	GGT	CG	GGG	GGCC	CGG	AG	CAC	CCC	AGG	AGT	ATAC	AC	AGG	TTCT	CA	1463	

y 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
b 1464 GCCTATCTCAACTGGATCTACAATGCTCGAAGGCTGAGCTG 1505

Search completed: November 23, 2003, 08:07:00
Job time : 280.081 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

Run on: November 23, 2003, 07:53:55 ; Search time 58.641 Seconds
(without alignments)
2197.846 Million cell updates/sec

Header: US-09-607-745-9
Perfect score: 1600
Sequence: 1 MDSKGSQKSRLLLLLVSN.....LNWIYVWKAELSRHHHHH 292

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

MODEL=frame+p2n.model -DEV=xlh
Q=/cgn2_1/USPTO_spool/US09607745/runat 21112003 144346 22244/app query.fasta_1.1038
DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
JUST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
USER=US09607745_@CGN_1_1_76 @runat 21112003 144346 22244 -NCPU=6 -ICPU=3
NO_MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -PGAPOP=6
XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1266	79.1	2038	3	US-09-008-271A-18
2	1266	79.1	2079	4	US-09-656-002-1
3	765	47.8	1130	4	US-09-387-375-8
4	754.5	47.2	1130	4	US-09-386-653A-8
5	746	46.6	1169	4	US-09-386-642-7
6	740.5	46.3	1166	4	US-09-386-629-2
7	672.5	42.0	1052	4	US-09-386-642-10
8	645	40.3	1049	4	US-09-386-642-9
9	601.5	37.6	1103	4	US-09-386-642-59
10	593	37.1	1142	4	US-09-386-642-8
11	574	35.9	1037	4	US-09-386-642-60
12	526.5	32.9	1460	4	US-09-370-838-80

13	526.5	32.9	1517	1	US-08-508-448C-15	Sequence 15, Appl
14	526.5	32.9	2790	4	US-09-370-838-79	Sequence 79, Appl
15	524.5	32.8	1462	4	US-09-370-838-55	Sequence 55, Appl
16	522.5	32.7	696	1	US-08-508-448C-24	Sequence 24, Appl
17	510.5	31.9	901	1	US-08-508-448C-9	Sequence 9, Appl
18	508	31.8	2479	3	US-09-342-749-29	Sequence 29, Appl
19	508	31.8	2479	4	US-09-691-840-29	Sequence 29, Appl
20	506	31.6	1479	3	US-09-342-749-1	Sequence 1, Appl
21	506	31.6	1479	4	US-09-691-840-1	Sequence 1, Appl
22	505.5	31.6	1615	4	US-09-820-002-1	Sequence 1, Appl
23	505.5	31.6	1783	3	US-09-510-738A-188	Sequence 188, App
24	505.5	31.6	1783	4	US-09-861-966-188	Sequence 188, App
25	505.5	31.6	2363	4	US-09-742-703-3	Sequence 3, Appl
26	502	31.4	2413	3	US-09-518-046-1	Sequence 1, Appl
27	497	31.1	1077	3	US-08-807-151-2	Sequence 2, Appl
28	497	31.1	1077	4	US-09-478-957-2	Sequence 2, Appl
29	491.5	30.7	1605	2	US-09-000-846-1	Sequence 1, Appl
30	482	30.1	959	4	US-09-023-942A-25	Sequence 25, Appl
31	478.5	29.9	2544	3	US-09-518-046-3	Sequence 3, Appl
32	473	29.6	2416	3	US-09-261-416-1	Sequence 1, Appl
33	456.5	28.5	1100	4	US-09-023-942A-5	Sequence 5, Appl
34	455	28.4	1613	4	US-09-387-375-1	Sequence 1, Appl
35	453.5	28.3	897	2	US-08-956-267A-1	Sequence 1, Appl
36	451.5	28.2	1081	3	US-09-008-271A-15	Sequence 15, Appl
37	450.5	28.2	1094	4	US-09-023-942A-3	Sequence 3, Appl
38	444	27.8	3147	2	US-09-027-337-1	Sequence 1, Appl
39	444	27.8	3147	4	US-09-644-600-1	Sequence 1, Appl
40	444	27.8	3147	4	US-09-644-600-18	Sequence 18, Appl
41	443.5	27.7	825	3	US-09-120-582-1	Sequence 1, Appl
42	442.5	27.7	980	4	US-09-023-942A-30	Sequence 30, Appl
43	442.5	27.7	1110	4	US-09-386-653A-1	Sequence 1, Appl
44	437.5	27.3	1212	4	US-09-620-312D-431	Sequence 431, App
45	431.5	27.0	1225	4	US-09-734-675-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-09-008-271A-18
; Sequence 18, Application US/09008271A
; Patent No. 6203979

GENERAL INFORMATION:

APPLICANT: Bandman, Olga
Hillman, Jennifer L.
Yue, Henry
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shah, Purvi

TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271A
FILING DATE: 16-Jan-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: COLNNOT13

CLONE: 1337018

SEQUENCE DESCRIPTION: SEQ ID NO: 18:

3-09-008-271A-18

Alignment Scores:
Pred. No.: 5.58e-130 Length: 2038
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
B: 3 Gaps: 0

S-09-607-745-9 (1-292) x US-09-008-271A-18 (1-2038)

Y	51	LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln	70
o	803	CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCCTGGCCCTGGCAGGTACGATCCAG	862
Y	71	TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla	90
o	863	TACGACAAACAGCACGTCTGTGGAGGAGCATCTGGACCCCACTGGTCCCTCAGGCA	922
Y	91	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	110
o	923	GCCCACTGCTTCAGGAAACATACCGATGTGTCACTGGAAGGTGGGGCAGGCTCAGAC	982
Y	111	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro	130
o	983	AAACTGGCAGCTCCCATCCCTGGCTGTGGCCAAAGATCATCATATTGAATCAACCCC	1042
Y	131	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	150
o	1043	ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCACTCCCACTCACTTCTCA	1102
Y	151	GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	170
b	1103	GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCC	1162
Y	171	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	190
o	1163	CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAAGATGTCTGACATACTG	1222
Y	191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	210
o	1223	CTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACCGGTGCAATGCAGACGATCGGTACCAG	1282
Y	211	GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	230
o	1283	GGGGAAGTCACCGAGAAGATGATGTGTGAGGATCCCGGAGGGGGTGTGGACACCTGC	1342
Y	231	GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	250
b	1343	CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGGCATC	1402
Y	251	ValSerTrpGlyTyrGlyCysGlyProSerThrProGlyValTyrThrLysValSer	270
b	1403	GTTAGCTGGGGCTATGGCTCGGGGGGCCGAGCACCCAGGAGTATACCAAGGTCTCA	1462
Y	271	AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	284
b	1463	GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG	1504

RESULT 2

US-09-656-002-1

; Sequence 1, Application US/09656002
; Patent No. 6455668

; GENERAL INFORMATION:

; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith

; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND

; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS

; FILE REFERENCE: A-69108/DJB/JJD/AMS

; CURRENT APPLICATION NUMBER: US/09/656,002

; CURRENT FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: US 09/525,993

; PRIOR FILING DATE: 2000-03-15

; PRIOR APPLICATION NUMBER: US 09/493,444

; PRIOR APPLICATION NUMBER: PCT/US 00/07044

; PRIOR FILING DATE: 2000-03-15

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1

; LENGTH: 2079

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-656-002-1

Alignment Scores:

Pred. No.:	5.74e-130	Length:	2079
Score:	1266.00	Matches:	228
Percent Similarity:	98.29%	Conservative:	2
Best Local Similarity:	97.44%	Mismatches:	4
Query Match:	79.12%	Indels:	0
DB:	4	Gaps:	0

US-09-607-745-9 (1-292) x US-09-656-002-1 (1-2079)

QY	51	LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln	70
Db	818	CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCCTGGCCCTGGCAGGTACGATCCAG	877
QY	71	TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla	90
Db	878	TACGACAAACAGCACGTCTGTGGAGGAGCATCTGGACCCCACTGGGTCTCAGGCA	937
QY	91	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	110
Db	938	GCCCACTGCTTCAGGAAACATACCGATGTGTCACTGGAAGGTGGGGCAGGCTCAGAC	997
QY	111	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro	130
Db	998	AAACTGGGCGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATATTGAATCAACCCC	1057
QY	131	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	150
Db	1058	ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCTCACTTCTCA	1117
QY	151	GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	170
Db	1118	GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCC	1177
QY	171	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	190
Db	1178	CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAAGATGTCTGACATACTG	1237
QY	191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	210
Db	1238	CTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACCGGTGCAATGCAGACGATCGGTACCAG	1297
QY	211	GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	230
Db	1298	GGGGAAGTCACCGAGAAGATGATGTGTGAGGAGTATCCCGAAGGGGTGTGGACACCTGC	1357


```
S-09-607-745-9 (1-292) x US-09-386-653A-8 (1-1130)
Y      1 MetAspSerLysGlySerSerGlnLysSerArgLeuLeuLeuValValSerAsn 20
b      13 ATGGACAGCAAGGTTGTCGCAGAAATCCCGCTGCTCCTGCTGGTGGTGTCAAAAT 72
Y      21 LeuLeuLeuCysGlnGlyValValSerAspTyrLysAspAspValAlaAla 40
b      73 CTACTCTTGCCAGGGTGTGCTCCGACTACAAGGACGACGACGCTGGACGCGGCC 132
Y      41 AlaLeuAlaAlaPropheAspAspAspLysIleValGlyGlyTyrAlaLeuAspVal 60
b      133 GCTCTTGCTGCCCTTTGATGATGATGACAAGATCGTTGGGGGCTATGCTCTAGAGGAG 192
Y      61 AspSerTrpProTrpGlnValSerIleGlnTyrAspLysGlnHisValCysGlyGlySer 80
b      193 GCGAGTGGCCCTGGCAAGTCAGCATCCAGCGCAACGGAAGCCACTTCTGCGGGGGCAGC 252
Y      81 IleLeuAspProHisTrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspVal 100
b      253 CTCATCGCGGAGCAGTGGGTCCTGACGGCTGCGCACTGCTTCCGCAACACCTCTGAGAGC 312
Y      101 PheAsnTrpLysValArgAlaGlySerAspLysLeu---GlySerPheProSerLeuAla 119
b      313 TCCCTGTACAGGTCCTGCTGGGGGCAAGGACAGCTAGTGACGCGGACACACGCTATG 372
Y      120 ValAlaLysIleIleIleGluPheAsnProMetTyr-----ProLysAspAsn 136
b      373 TATGCCCGGTGAGGAGGAGTGAGAGCAACCCCTGTACCAGGGCAGCGCCTCCAGCGCT 432
Y      137 AspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIle 156
b      433 GACGTGGCCCTGTGGAGCTGGAGGACACGAGTGCCTTCACCAATTACATCCTCCCGCTG 492
Y      157 CysLeuProPhePheAspGluGluLeuThrProAlaThrProLeu-----TrpIleIle 174
b      493 TGCCTGCCT-----GACCCCTCGGTGATCTTTGAGACGGGCGATGAAGTCTGGGTCACT 546
Y      175 GlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAsp----- 188
b      547 GGCTGGGGCAGCCCGCAGTGAG-----GAAGACCTCCTGCCGGAACCGCGG 591
Y      189 IleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsn----- 204
b      592 ATCTGTCAGAAACTCGTGTGCCCATCATCGACACACCCCAAGTGAACCTGTCTCTACAGC 651
Y      205 -----AlaAspAspAlaTyrGlnGlyGlu---ValThrGluLysMetMetCysAlaGly 221
b      652 AAAGACACCGAGTTGGCTACCAACCCAAACCATCAAGAAATGACATGCTGTGCGCGCGC 711
Y      222 IleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGln 241
b      712 TTCGAGGAGGGCAAGAAGGATGCCTGCAAGGGCGACTCGGGCGGCCCTGTGTGCGCTC 771
Y      242 SerAspGln---TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyPro 260
b      772 GTGGGTCACTCGTGGCTGCAGCGGGGGGTGATCAGCTGGGGGTGAGGGCTGTGCCCGCCAG 831
Y      261 SerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnVal--- 279
b      832 AACCGCCAGGTGTCTACATCCGTGTACCGGCCCCACCAACTGGATCCATCGGATCATC 891
Y      280 -----TrpLysAlaGluLeuSerArgHisHisHisHis 290
b      892 CCCAAACTGCAGTTCAGCCAGCGAGGTTGGGGCGGCCAGAAAGTCTAGACATCACCATCAC 951
Y      291 HisHis 292
b      952 CATCAC 957
```

RESULT 5

S-09-386-642-7

Sequence 7, Application US/09386642

```
Patent No. 6420157
GENERAL INFORMATION:
APPLICANT: Darrow, Andrew
APPLICANT: Qi, Jensen
APPLICANT: Andrade-Gordon, Patricia
TITLE OF INVENTION: Zymogen Activation System
FILE REFERENCE: ORT-1028
CURRENT APPLICATION NUMBER: US/09/386,642
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 1169
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
OTHER INFORMATION: with homo sapien serine protease catalytic domain
US-09-386-642-7
```

```
Alignment Scores:
Pred. No.: 5.83e-73 Length: 1169
Score: 746.00 Matches: 163
Percent Similarity: 60.37% Conservative: 35
Best Local Similarity: 49.70% Mismatches: 94
Query Match: 46.62% Indels: 36
DB: 4 Gaps: 8
```

US-09-607-745-9 (1-292) x US-09-386-642-7 (1-1169)

```
QY      1 MetAspSerLysGlySerSerGlnLysSerArgLeuLeuLeuValValSerAsn 20
DB      13 ATGGACAGCAAGGTTGTCGCAGAAATCCCGCTGCTCCTGCTGGTGGTGTCAAAAT 72
QY      21 LeuLeuLeuCysGlnGlyValValSerAspTyrLysAspAspValAlaAla 40
DB      73 CTACTCTTGTCGCGGGTGTGGTCTCCGACTACAAGGACGACGACGCTGGACGCGGCC 132
QY      41 AlaLeuAlaAlaPropheAspAspAspLysIleValGlyGlyTyrAlaLeuAspVal 60
DB      133 GCTCTTGCTGCCCTTTGATGATGATGACAAGATCGTTGGGGGCTATGCTCTAGAGGCC 192
QY      61 AspSerTrpProTrpGlnValSerIleGlnTyrAspLysGlnHisValCysGlyGlySer 80
DB      193 GGTCACTGGCCCTGGCAGGTGAGCATCACCTATGAAGCGCTCCATGTGTGTGGTGGCTCT 252
QY      81 IleLeuAspProHisTrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspVal 100
DB      253 CTCGTGTGTGAGCAGTGGGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 312
QY      101 PheAsnTrpLysValArgAlaGlySerAspLysLeuGlySerPheProSerLeuAla--- 119
DB      313 GAAGCCTATGAGGTCAAGCTGGGGGCCCCACGACTAGACTCTACTCCGAGGACGCGCAAG 372
QY      120 ValAlaLysIleIleIleGluPheAsnProMetTyrProLysAsp-----Asn 136
DB      373 GTCAGCACCTGTAAGGACATCATCCCCACCCAGTACCTCCAGAGGGTCCCGAGGGC 432
QY      137 AspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIle 156
DB      433 GACATTGCACCTCTCCAACTCAGCAGACCCCATCCTTCTCCCGTACATCCCGGCCCATC 492
QY      157 CysLeuProPhePheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrp 176
DB      493 TGCCTCCTGTCAGCCCAACGCTCCTTCCCCAACGCGCTCCACTGCACTGTCTGCTGGCTGG 552
QY      177 GlyPheThrLysGlnAsnGlyGlyLysMetSer---AspIleLeuLeuGlnAlaSerVal 195
DB      553 GGTCTGTGGCCCCCTCAGTGAGCCCTCCTGACGCGCCCAAGCCACTGCAGCAACTCGAGGTG 612
QY      196 GlnValIleAspSerThrArgCysAsnAla-----AspAspAlaTyrGlnGlyGlu 212
DB      613 CCTCTGATCAGTGTGAGAGCGTGTAACTGCCTGTACAACATCGACGCCCAAGCCTGAGGAG 672
```


OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
OTHER INFORMATION: with homo sapien serine protease catalytic domain

Alignment Scores:		6.43e-65	Length:	1052
Score:	672.50	Matches:	148	
Percent Similarity:	61.31%	Conservative:	39	
Best Local Similarity:	48.52%	Mismatches:	89	
Query Match:	42.03%	Indels:	29	
DB:	4	Gaps:	11	

US-09-607-745-9 (1-292) x US-09-386-642-10 (1-1052)

Y	1	MetAspSerLysGlySerSerGlnLysSerArgLeuLeuLeuLeuValValSerAsn	20
b	13	ATGGACAGCAAGGTTGTCGCAGAAATCCCGCTGCTCTGCTGCTGGTGTGTTCAAT	72
Y	21	LeuLeuLeuCysGlnGlyValValSerAspTyrLysAspAspValAlaAla	40
b	73	CTACTCTTGTGCAGGGTGTGGTCTCCGACTACAGGACGACGACGTGGACGGCC	132
Y	41	AlaLeuAlaAlaPropheAspAspAspLysIleValGlyTyr--AlaLeuAsp	59
b	133	GCTCTTGTGTCCTTGTATGATGATGACAAAGATCGTTGGGGCTACAACTGTCTAGAA	192
Y	60	ValAspSerTrpProTrpGlnValSerIleGlnTyrAspLysGlnHisValCysGly	79
b	193	AAGCACTCCAGCCCTGGCAGGCGAGCCCTGTCGAGAAGACGGGCTACTCTGTGGGGCG	252
Y	80	SerIleLeuAspProHisTrpValLeuThrAlaAlaHisCysPheArgLys	96
b	253	ACGCTCATCGCCCGCCAGATGGTCTCTGACAGAGCCCTCAAGCCCGCTACATA	312
Y	97	--HisThrAspValPheAsnTrpLysValArgAlaGlySerAspLys	111
b	313	GTTACCTGGGCGCAGCACACCTCCAGAGGAGGAGGCTGTGACGAGACCCGGACAGCC	372
Y	112	LeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnProMet	131
b	373	ACTGAGTCTTCCCCACCCCGGC-----TTCAACAACAGC	408
Y	132	TyrPro---LysAsp-----AsnAspIleAlaLeuMetLysLeuGlnPheProLeuThr	148
b	409	CTCCCCAACAAAGACCACCGCAATGACATCATGCTGGTGAAGATGGCATCGCCAGTCTCC	468
Y	149	PheSerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAla	168
b	469	ATCACCTGGGCTGTGCGACCCCTCACCTCTCC-----TCACGCTGTGTCACTGTGGC	522
Y	169	ThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAsp	188
b	523	ACCAGCTGCTCATTTCCGGCTGGGGCAGCACGTCCAGCCCCCAGTTACGCTGCTCAC	582
Y	189	IleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAla	208
b	583	ACCTTGGGATGCGCAACATCACCATCATTTGACACCCAGAGTGT-----GAGAAGCC	636
Y	209	TyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAsp	228
b	637	TACCCCGGAACATCACAGACACCATGGTGTGTGCCAGGTGCAGGAAGGGGCAAGGAC	696
Y	229	ThrCysGlnGlyAspSerGlyProLeuMetTyrGlnSerAspGlnTrpHisValVal	248
b	697	TCCTGCCAGGTGACTCCGGGGGCCCTCTGGTCTGTAAACAGTCT-----CTTCAA	747
Y	249	GlyIleValSerTrpGlyTyrGly---CysGlyGlyProSerThrProGlyValTyrThr	267
b	748	GGCATTATCTCTGGGGCCAGGATCCGTGTGCGATCACCCGAAAGCCTGGTGTCTACAG	807
Y	268	LysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeuSerArgHis	287
b	808	AAAGTCTGCAATATGTGGACTGGATCCAGGAGACGATGAAGACAAT---TCTAGACAT	864

QY	288	HishHishHishHis	292
Db	865	CACCATCACCATCAC	879

RESULT 8
US-09-386-642-9
; Sequence 9, Application US/09386642
; Patent No. 6420157
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jensen
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Zymogen Activation System
; FILE REFERENCE: ORT-1028
; CURRENT APPLICATION NUMBER: US/09/386,642
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1049
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
; OTHER INFORMATION: with homo sapien serine protease catalytic domain
US-09-386-642-9

Alignment Scores:
Pred. No.: 6.92e-62 Length: 1049
Score: 645.00 Matches: 146
Percent Similarity: 60.33% Conservative: 38
Best Local Similarity: 47.87% Mismatches: 91
Query Match: 40.31% Indels: 30
DB: 4 Gaps: 11

US-09-607-745-9 (1-292) x US-09-386-642-9 (1-1049)	
QY	1 MetAspSerLysGlySerSerGlnLysSerArgLeuLeuLeuValValSerAsn 20
Db	13 ATGGACAGCAAGGTTGTCGCAGAAATCCCGCTGCTCTGCTGTGTGTCTCAAAAT 72
QY	21 LeuLeuLeuCysGlnGlyValValSerAspTyrLysAspAspValAspAlaAla 40
Db	73 CTACTCTTGTGCCAGGTTGTTCTCCGACTACAGGACGACGACGCTGGACGGGCC 132
QY	41 AlaLeuAlaAlaPropheAspAspAspLysIleValGlyGlyTyr---AlaLeuAsp 59
Db	133 GCTCTTGTGCTGCCCTTTGATGATGATGACAGATCGTTGGGGCTACAACTGTCTAGAA 192
QY	60 ValAspSerTrpProTrpGlnValSerIleGlnTyrAspLysGlnHisValCysGlyGly 79
Db	193 CCCCATTCGACGCTTGGCAGGGCGGCTTGTTCAGGGCCAGCAACTACTCTGTGGCGGT 252
QY	80 SerIleLeuAspProHisTrpValLeuThrAlaAlaHisCysPheArg---LysHisThr 98
Db	253 GTCCCTGTAGGTGGCAACTGGGTCTTTACAGTGGCCACTGTAAAAAACCGAAATACACA 312
QY	99 AspValPheAsnTrpLysValArgAlaGlySerAspLysLeu-----GlySer 114
Db	313 -----GTACGCTGGGAGACACAGCCTACAGATAAAGATGGCCCA 354
QY	115 PheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnProMetTyr----- 132
Db	355 GAGCAAGAAATACCTGTGGTTCAGTCCATCCCA-----CACCCCTGTACACAGC 405
QY	133 -----ProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThr 148
Db	406 AGCGATGTGGAGGACCAACCATGATCTGATGCTTCTTCAACTGCGTGACCGAGGCATCC 465
QY	149 PheSerGlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAla 168
Db	466 CTGGGGTCCAAAGTGAAGCCCATCAGCCTGGCA-----GATCATTCACCCAGCCTGGC 519

169 ThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAsp 188
520 CAGAAAGTGCACCGTCTCAGGCTGGGCACTGTCCAGTCCCGAGAGAAATTTTCCTGAC 579
189 IleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAla 208
580 ACTCTCAACTGTGCAGAGTAAATCTTCCCGAGAAGAGTGT-----GAGGATGCT 633
209 TyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAsp 228
634 TACCCGGGCGAGATCACAGATGGCATGGTCTGTGCAGGCAGCAGCAA---GGGGCTGAC 690
229 ThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValVal 248
691 ACGTGCCAGGGGATCTTGGAGGCCCTGGTGTGTGATGGT-----GCATCCAG 741
249 GlyIleValSerTrpGlyTyrGly---CysGlyGlyProSerThrProGlyValTyrThr 267
742 GGCATCACATCTGGGGCTCAGACCCCTGTGGAGGTCCGACAAACCTGGCGTCTATACC 801
268 LysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeuSerArgHis 287
802 AACATCTGCCGCTACCTGGACTGGATCAAGAAGATCATAGGCAGCAAGGGCTCTAGACAT 861
288 HisHisHisHis 292
862 CACCATCACCATCAC 876

RESULT 9

S-09-386-642-59
Sequence 59, Application US/09386642
Patent No. 6420157

GENERAL INFORMATION:

APPLICANT: Darrow, Andrew
APPLICANT: Qi, Jensen
APPLICANT: Andrade-Gordon, Patricia
TITLE OF INVENTION: Zymogen Activation System
FILE REFERENCE: ORT-1028
CURRENT APPLICATION NUMBER: US/09/386,642
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 59

LENGTH: 1103
TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Nucleic acid
OTHER INFORMATION: sequence of human protease F in CFEK2 zymogen
OTHER INFORMATION: vector

S-09-386-642-59

Alignment Scores:

red. No.:	4.67e-57	Length:	1103
core:	601.50	Matches:	126
percent Similarity:	58.50%	Conservative:	46
est local Similarity:	42.86%	Mismatches:	85
very Match:	37.59%	Indels:	37
3:	4	Gaps:	10

S-09-607-745-9 (1-292) x US-09-386-642-59 (1-1103)

28 ValSerAspTyrLysAspAspValAspAlaAlaLeuAlaAlaProPheAsp 47
73 GTCCCCGACTACAGACGACGAC-----GACGCGCGCTCTGCTGCCCTTTGAT 126
48 AspAspAspLysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnVal 67
127 GATGATGACAGATCGTTGGGGCTATGCTCTAGAACTCGGGCGCTGGCCGTGGCAGGG 186
68 SerIleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpVal 87

Db 187 AGCCTGGCCTGTGGATTCCACGATATCGGAGTGGAGCTGCTCAGCCACCGCTGGGCA 246
QY 88 LeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsn-----TrpLys 104
Db 247 CTCACGGCGCGCACTGCTTTGAAACCTATAGTAGACCTTAGTGATCCCTCCGGGTGGATG 306
QY 105 ValArgAlaGlySerAspLysLeuGlySerPhePheProSerLeuAlaValAlaLys----- 122
Db 307 GTCCAGTTTGGC-----CAGTGACTTCCATGCCATCCTTCTGGAGCCTGCAGGCTAC 360
QY 123 -----IleIleIleIleGluPheAsnProMetTyrProLysAspAsn--- 136
Db 361 TACAACCGTTACTTCGTATCGAATATCTATCTGAGCCCTCGCTACCTGGGAATTCACCC 420
QY 137 ---AspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArgPro 155
Db 421-TATGACATTCCTTGGTGAAGTGTCTGCACCTGTCTACCTTACACTAAACACATCCAGCCC 480
QY 156 IleCysLeuProPhePheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGly 175
Db 481 ATCTGTCTCCAGGCCCTCCACATTGAGTTTGAGAAACCGGACAGACTGTCTGGTGACTGGC 540
QY 176 TrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeuLeuGln---AlaSer 194
Db 541 TGGGGGTACATCAAAGAGGATGAGGCACTGCCATCTCCCCACACCCCTCCAGGAAGTTCAG 600
QY 195 ValGlnValIleAspSerThrArgCysAsn-----AlaAspAlaTyrGlnGly 211
Db 601 GTCGCCATCATAAAACAACCTCTATGTGCAACCACCTCTTCTCAAGTACAGTTTCCGCAAG 660
QY 212 GluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCysGln 231
Db 661 GACATCTTTGGAGACATGTTTGTGTGCTGGCAATGCCCAAGCGGGAAGGATGCCTGCTTC 720
QY 232 GlyAspSerGlyGlyProLeu---MetTyrGlnSerAspGlnTrpHisValValGlyIle 250
Db 721 GGTGACTCAGGTGGACCCCTTGGCCTGTAAACAAGATGGACTGTGGTATCAGATTGGAGTC 780
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
Db 781 GTGAGCTGGGAGTGGGCTGTGTGGCCCAATCGGCCCGGTGTCTACACCAATATCAGC 840
QY 271 AlaTyrLeuAsnTrpIleTyrAsnVal----- 279
Db 841 CACCACCTTGAGTGGATCCAGAAGCTGATGGCCAGAGTGGCATGTCCCAGCCAGACCCC 900
QY 280 ---TrpLysAlaGluLeuSerArgHisHisHisHisHis 292
Db 901 TCCTGG-----TCTAGACATCACCATCACCATCAC 930

RESULT 10

US-09-386-642-8

; Sequence 8, Application US/09386642
; Patent No. 6420157

GENERAL INFORMATION:

APPLICANT: Darrow, Andrew
APPLICANT: Qi, Jensen
APPLICANT: Andrade-Gordon, Patricia
TITLE OF INVENTION: Zymogen Activation System
FILE REFERENCE: ORT-1028
CURRENT APPLICATION NUMBER: US/09/386,642
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 1142
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
OTHER INFORMATION: with homo sapien serine protease catalytic domain

US-09-386-642-8

532 GGCTGGGGTCTGCTGGCG-----AACGGGAGAAATGCCTACCGTGTGCTGAGTGGTGAAC 585
195 ValGlnValIleAspSerThrArgCysAsnAla-----AspAspAlaTyrGlnGlyGlu 212
586 GTGTCGGTGGTCTGAGGAGGTCTGCAGTAAGCTCTATGACCCGCTGTACCAC----- 639
213 ValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCysGlnGly 232
640 -----CCCAGCATGTTCTGCGCGGGGAGGACGACGACGAGGACTCCTGCAACGGT 693
233 AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIleValSer 252
694 GACTCTGGGGGCCCCCTGATCTGCAACGGG-----TACTGCAAGGCTTGTGTCT 744
253 TrpGlyTyrGly---CysGlyGlyProSerThrProGlyValTyrThrLysValSerAla 271
745 TTCGGAAAGCCCCGTGTGGCCAAAGTTGGCGTGCCAGGTGTCTACACCAACCTCTGCAAA 804
272 TyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeuSerArgHisHisHisHis 291
805 TTCACGTAGTGGATAGAGAAACCGTCCAGGCCAGT---TCTAGACATCACCATCACCAT 861
292 His 292
862 CAC 864

RESULT 12

3-09-370-838-80
Sequence 80, Application US/09370838
Patent No. 644425
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Mohamath, Roadoh
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.475C1
CURRENT APPLICATION NUMBER: US/09/370,838
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/285,323
EARLIER FILING DATE: 1999-04-02
NUMBER OF SEQ ID NOS: 289
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 80
LENGTH: 1460
TYPE: DNA
ORGANISM: Homo sapien
-09-370-838-80

ignment Scores:
ed. No.: 1.31e-48 Length: 1460
ore: 526.50 Matches: 97
rcent Similarity: 62.93% Conservative: 49
st Local Similarity: 41.81% Mismatches: 79
ery Match: 32.91% Indels: 7
4 Gaps: 4
-09-607-745-9 (1-292) x US-09-370-838-80 (1-1460)

49 AspAspLysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSer 68
601 GAGCAGAGATCCTTGGAGGCACTGAGGCTGAGGAGGAAAGCTGGCCGIGGCAAGTCAGT 660
69 IleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu 88
661 CTGCGGCTCAATAATGCCCACTGTGGAGGCACTGATCAATAACATGTGGATCCTG 720
89 ThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGly 108
721 ACAGCAGCTCACTGCTCAGAAGCAACTCTATCTCTGCTGACTGGATTGCCAGCTCTGGT 780
109 SerAspLysLeuGlySerPheProSerLeu-----AlaValAlaLysIleIleIle 126

Db 781 ATTTCCACA-----ACATTTCTAAACTAAGATGAGAGTAAGAAATATTTTAATTCAT 834
QY 127 GluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhePro 146
Db 835 AACAAATTATAAATCTGCAACTCATGAAATGACATTCGACTTGTGAGACTTGAGAACAGT 894
QY 147 LeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThr 166
Db 895 GTCACCTTTACCAAGATATCCATAGTGTGTCTCCAGCTGTACCCAGAAATATTCGA 954
QY 167 ProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMet 186
Db 955 CCTGGCTCTACTGCTTATGTAAACAGGATGGGCGCTCAAGAAATATGCTGGCCACACAGT 1014
QY 187 SerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAsp 206
Db 1015 CCAGAG---CTAAGGCAAGGACAGGTGAGAAATAATAAGTAATGATGTATGTATGCACCA 1071
QY 207 AspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGly 226
Db 1072 CATAGTTATAATGGAGCCATCTGTCTGGAATGCTGTGTGCTGGAGTACCTCAAGGTGGA 1131
QY 227 ValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGln----- 244
Db 1132 GTGGACCATGTTCAGGGTGTACTCTGGTGGCCCACTAGTACAAGAGACTCACGGCGGCTT 1191
QY 245 TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGly 264
Db 1192 TGGTTATTGTGGGATAGTAAGCTGGGAGATCAGTGTGGCTGCCGGATAAGCCAGGA 1251
QY 265 ValTyrThrLysValSerAlaTyrLeuAsnTrpIle 276
Db 1252 GTGTACTCGAGTGACAGCCTACCTTGTACTGGATT 1287

RESULT 13

US-08-508-448C-15
; Sequence 15, Application US/08508448C
; Patent No. 5804410
; GENERAL INFORMATION:
; APPLICANT: Kazuyoshi YAMAOKA et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
; TITLE OF INVENTION: TRYPsin-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,448C
; FILING DATE: July 28, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:

265 ValTyrThrLysValSerAlaTyrLeuAsnTrpIle 276
1248 GTGTATACTCGAGTGACAGCCTACCTTGACTGGATT 1283

SULT 15
-09-370-838-55
Sequence 55, Application US/09370838
Patent No. 644425
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Mohamath, Roadoh
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.475C1
CURRENT APPLICATION NUMBER: US/09/370,838
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/285,323
EARLIER FILING DATE: 1999-04-02
NUMBER OF SEQ ID NOS: 289
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 55
LENGTH: 1462
TYPE: DNA
ORGANISM: Homo sapien
S-09-370-838-55

Alignment Scores:
red. No.: 2.17e-48 Length: 1462
core: 524.50 Matches: 96
percent Similarity: 62.93% Conservative: 50
best Local Similarity: 41.38% Mismatches: 79
every Match: 32.78% Indels: 7
3: 4 Gaps: 4

S-09-607-745-9 (1-292) x US-09-370-838-55 (1-1462)
49 AspAspLysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSer 68
597 GAGCAGAGAAATCCTTGGAGGCACTGAGGCTGAGGAGGAGCTGGCCGTGGCAAGTCAGT 656
69 IleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu 88
657 CTGGGGCTCAATAATGCCACCACCTGTGGAGGCAGCCTGATCAATAACATGTGGATCCTG 716
89 ThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGly 108
717 ACAGCAGCTCACTGCTTCAGAAAGCAACTTAATCCTCGTGAAGTGGATTGCCACGCTCGGT 776
109 SerAspLysLeuGlySerPheProSerLeu-----AlaValAlaLysIleIleIle 126
777 ATTTCCACA-----ACATTTCCTAACTAAGATGAGAGTAAGAAATATTTTAATTCAT 830
127 GluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhePro 146
831 AACAAATTATAATCTGCAACTCATGAAAATGACATTGCACCTTGTGAGACTTGAGAACAGT 890
147 LeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThr 166
891 GTCACCTTTACCAAGATATCCATAGTGTGTGTCTCCACGCTGCTACCCAGAAATATCCA 950
167 ProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMet 186
951 CCTGGCTCTACTGCTTATGTAAACAGGATGGGCGCTCAAGAATATGCTGGCCACACAGTT 1010
187 SerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAsp 206
1011 CCAGAG--CTAAGGCAAGGACAGGTCAGATAATAAGTAATGATGTATGTAATGCACCA 1067
207 AspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGly 226
1068 CATAGTTATAATGGAGCCATCTTTGCTGGAATGCTGTGTGTGGAGTACCTCAAGGTGGA 1127

QY 227 ValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGln----- 244
Db 1128 GTGGACGCATGTTCAGGGTGAATCTGGTGGCCCACTAGTACAAGAAGACTCACGGCGGCTT 1187
QY 245 TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGly 264
Db 1188 TGGTTTATTGTGGGATAGTAAGCTGGGGAGATCATCATGTGGCCTGCCGGATAAGCCAGGA 1247
QY 265 ValTyrThrLysValSerAlaTyrLeuAsnTrpIle 276
Db 1248 GTGTATACTCGAGTGACAGCATACATTGACTGGATT 1283

Search completed: November 23, 2003, 12:09:06
Job time : 71.641 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.
M protein - nucleic search, using frame_plus_p2n model
run on: November 23, 2003, 07:54:55 ; Search time 314.091 Seconds
(without alignments)
3038.343 Million cell updates/sec

title: US-09-607-745-9
perform score: 1600
sequence: 1 MDSKSSQKSRLLLLLVVSN.....LNWIYVWKAELSRHHHHH 292

scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

searched: 2169961 seqs, 1634102185 residues

total number of hits satisfying chosen parameters: 4339922

minimum DB seq length: 0
maximum DB seq length: 2000000000
post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

command line parameters: MODEL=frame+ p2n.model -DEV=xlh
Q=/cgn2_1/USPTO_spool/US09607745/runat_21112003_144347_22270/app_query.fasta_1.1038
DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
MAXLEN=200000000 -USER=US09607745 @CGN 1 1 424 @runat 21112003_144347_22270
NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
JONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
XGAPOP=6 -XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

database : Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Match	Length	DB ID	Description
1	1266	79.1	2038	12	US-10-180-719-18
Sequence 18, Appl					

2	1266	79.1	2063	11	US-09-888-257A-2	Sequence 2, Appli
3	1266	79.1	2063	11	US-09-946-374-274	Sequence 274, App
4	1266	79.1	2063	12	US-10-015-387A-274	Sequence 274, App
5	1266	79.1	2063	12	US-10-063-735-111	Sequence 111, App
6	1266	79.1	2063	12	US-10-006-130A-274	Sequence 274, App
7	1266	79.1	2063	12	US-10-199-672-329	Sequence 329, App
8	1266	79.1	2063	12	US-10-006-172A-274	Sequence 274, App
9	1266	79.1	2063	12	US-10-187-749-329	Sequence 329, App
10	1266	79.1	2063	12	US-10-194-457-329	Sequence 329, App
11	1266	79.1	2063	12	US-10-184-642-329	Sequence 329, App
12	1266	79.1	2063	12	US-10-196-747-329	Sequence 329, App
13	1266	79.1	2063	12	US-10-015-392A-274	Sequence 274, App
14	1266	79.1	2063	12	US-10-017-253A-274	Sequence 274, App
15	1266	79.1	2063	12	US-10-173-689-329	Sequence 329, App
16	1266	79.1	2063	12	US-10-173-690-329	Sequence 329, App
17	1266	79.1	2063	12	US-10-173-691-329	Sequence 329, App
18	1266	79.1	2063	12	US-10-173-692-329	Sequence 329, App
19	1266	79.1	2063	12	US-10-173-694-329	Sequence 329, App
20	1266	79.1	2063	12	US-10-173-698-329	Sequence 329, App
21	1266	79.1	2063	12	US-10-173-699-329	Sequence 329, App
22	1266	79.1	2063	12	US-10-173-707-329	Sequence 329, App
23	1266	79.1	2063	12	US-10-174-569-329	Sequence 329, App
24	1266	79.1	2063	12	US-10-174-583-329	Sequence 329, App
25	1266	79.1	2063	12	US-10-174-587-329	Sequence 329, App
26	1266	79.1	2063	12	US-10-174-589-329	Sequence 329, App
27	1266	79.1	2063	12	US-10-174-591-329	Sequence 329, App
28	1266	79.1	2063	12	US-10-175-736-329	Sequence 329, App
29	1266	79.1	2063	12	US-10-175-742-329	Sequence 329, App
30	1266	79.1	2063	12	US-10-175-744-329	Sequence 329, App
31	1266	79.1	2063	12	US-10-175-745-329	Sequence 329, App
32	1266	79.1	2063	12	US-10-175-748-329	Sequence 329, App
33	1266	79.1	2063	12	US-10-175-751-329	Sequence 329, App
34	1266	79.1	2063	12	US-10-175-754-329	Sequence 329, App
35	1266	79.1	2063	12	US-10-176-480-329	Sequence 329, App
36	1266	79.1	2063	12	US-10-176-489-329	Sequence 329, App
37	1266	79.1	2063	12	US-10-176-754-329	Sequence 329, App
38	1266	79.1	2063	12	US-10-176-755-329	Sequence 329, App
39	1266	79.1	2063	12	US-10-176-759-329	Sequence 329, App
40	1266	79.1	2063	12	US-10-176-920-329	Sequence 329, App
41	1266	79.1	2063	12	US-10-176-922-329	Sequence 329, App
42	1266	79.1	2063	12	US-10-176-924-329	Sequence 329, App
43	1266	79.1	2063	12	US-10-176-984-329	Sequence 329, App
44	1266	79.1	2063	12	US-10-179-508-329	Sequence 329, App
45	1266	79.1	2063	12	US-10-179-512-329	Sequence 329, App

ALIGNMENTS

RESULT 1
US-10-180-719-18
; Sequence 18, Application US/10180719
; Publication No. US20030166246A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/180,719
FILING DATE: 25-Jun-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271
FILING DATE: 16-Jan-1998
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNNOT13
CLONE: 1337018

SEQUENCE DESCRIPTION: SEQ ID NO: 18:
S-10-180-719-18

Alignment Scores:
red. No.: 1.58e-152 Length: 2038
core: 1266.00 Matches: 228
percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
B: 12 Gaps: 0

S-09-607-745-9 (1-292) x US-10-180-719-18 (1-2038)

Y 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
: : : : :
D 803 CGTGTGGTGGTGGGAGGAGGCGCTCTGTGGATTCTTGGCCTGGCAGGTCAGCATCCAG 862
Y 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
: : : : :
D 863 TAGCAAAACAGCACGCTCTGTGGAGGAGCATCTGGACCCCACTGGTCTCCTCAGGCA 922
Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
: : : : :
D 923 GCCACACTGCTCAGGAAACATACCGATGTGTTCACTGGAAGGTGGGGCAGGTCAGAC 982
Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
: : : : :
D 983 AAACGTGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATATTGAATTCAACCCC 1042
Y 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
: : : : :
D 1043 ATGTACCCCAAGACATGACATCGCCCTCATGAAGCTGAGTCCAGTCCCACTCACTTTCTCA 1102
Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAlaThrPro 170
: : : : :
D 1103 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1162
Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu 190
: : : : :
D 1163 CTCTGGATCATGTGATGGGCTTTACGAAGCAGATGGAGGAAGATGCTGTACATACTG 1222
Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 210
: : : : :
D 1223 CTGCAGGCGTCAGTCCAGGTTCATTGACAGCACCGGTGCAATGCAGACGATGCTGACACTG 1282
Y 211 GlyGlnValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
: : : : :
D 1283 GGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGAAGGGGGTGTGGACACCTGC 1342
Y 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250

Db 1343 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGGCATC 1402
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
: : : : :
Db 1403 GTTAGCTGGGGCTATGGCTGGGGGGCCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1462
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
: : : : :
Db 1463 GCCTATCTCAACTGGATCTACATATGTCTGGAAGGCTGAGCTG 1504

RESULT 2

US-09-888-257A-2
; Sequence 2, Application US/09888257A
; Publication No. US20030060612A1
; GENERAL INFORMATION:
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Polakis, Paul
; APPLICANT: Smith, Victoria
; APPLICANT: Wood, William I.
; APPLICANT: Wu, Thomas D.
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF TUMOR

; FILE REFERENCE: P5002R1

; CURRENT APPLICATION NUMBER: US/09/888,257A

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/063,540

; PRIOR FILING DATE: 1997-10-28

; PRIOR APPLICATION NUMBER: US 60/089,653

; PRIOR FILING DATE: 1998-06-17

; PRIOR APPLICATION NUMBER: US 60/099,792

; PRIOR FILING DATE: 1998-09-10

; PRIOR APPLICATION NUMBER: US 60/103,678

; PRIOR FILING DATE: 1998-10-08

; PRIOR APPLICATION NUMBER: US 60/235,451

; PRIOR FILING DATE: 2000-09-26

; PRIOR APPLICATION NUMBER: PCT/US99/12252

; PRIOR FILING DATE: 1999-06-02

; PRIOR APPLICATION NUMBER: PCT/US99/20111

; PRIOR FILING DATE: 1999-09-01

; PRIOR APPLICATION NUMBER: PCT/US00/04342

; PRIOR FILING DATE: 2000-02-18

; PRIOR APPLICATION NUMBER: PCT/US00/05841

; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: PCT/US00/08439

; PRIOR FILING DATE: 2000-03-30

; PRIOR APPLICATION NUMBER: PCT/US00/23328

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: PCT/US00/32678

; PRIOR FILING DATE: 2000-12-01

; PRIOR APPLICATION NUMBER: PCT/US01/06520

; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: PCT/US01/06666

; PRIOR FILING DATE: 2001-03-01

; NUMBER OF SEQ ID NOS: 10

; SEQ ID NO 2

; LENGTH: 2063

; TYPE: DNA

; ORGANISM: Homo Sapien

US-09-888-257A-2

Alignment Scores:

Pred. No.:

Score: 1.6e-152 Length: 2063

Percent Similarity: 1266.00 Matches: 228

Best Local Similarity: 98.29% Conservative: 2

Query Match: 97.44% Mismatches: 4

Indels: 79.12% Indels: 0

DB: 11 Gaps: 0

-09-607-745-9 (1-292) x US-09-888-257A-2 (1-2063)

51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
: : : : :
804 CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTCAGCATCCAG 863

71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
: : : : :
864 TACGACAAACAGCACGCTCTGTGGAGGGAGCATCCTGGACCCCACTGGTCTCTCAGGCAC 923

91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
: : : : :
924 GCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCAGAC 983

111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
: : : : :
984 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATCATTAATCAACCCC 1043

131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
: : : : :
1044 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGTGCAGTTCCTCCACTCTCTCTCA 1103

151 GlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAlaThrPro 170
: : : : :
1104 GGCACAGTCAGGCCCATCTGTGTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163

171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
: : : : :
1164 CTCTGGATCATTTGGATGGGGTTTACGAAGCAGAAATGAGGGAAGATGTCTGACATACTG 1223

191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 210
: : : : :
1224 CTGCAGGCGTCAGTCCAGGTCATGTGACAGCACACGGTGCATGCAGACGATCGGTACGAC 1283

211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
: : : : :
1284 GGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGGAAGGGGTGTGGACACCTGC 1343

231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
: : : : :
1344 CAGGGTGACAGTGGTGGGCCCTGTATGTACCAATCTGACCAGTGGCATGTGGTGGGCATC 1403

251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
: : : : :
1404 GTTAGCTGGGCTATGGCTCGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1463

271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
: : : : :
1464 GCCTATCTCAACTGGAATCAATGTCTGGAAGGCTGAGCTG 1505

SULT 3

-09-946-374-274
Sequence 274, Application US/09946374
Publication No. US20030073129A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.

APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth J.

APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann

APPLICANT: Smith, Victoria

APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel

APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
ACIDS OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2830P1C1
CURRENT APPLICATION NUMBER: US/09/946,374
CURRENT FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 60/098716
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098723
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098749
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098750
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098803
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098821
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098843
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/099536
PRIOR FILING DATE: 1998-09-03
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-03
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-03
PRIOR APPLICATION NUMBER: 60/099602
PRIOR FILING DATE: 1998-09-03
PRIOR APPLICATION NUMBER: 60/099642
PRIOR FILING DATE: 1998-09-03
PRIOR APPLICATION NUMBER: 60/099741
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099754
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099763
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099792
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099808
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099815
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100388
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100584
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100661
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100662
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100664
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100683
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100684
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100710
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100711
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100848

PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100849
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100930
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101014
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101068
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101071
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101279
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: 60/101471
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101472
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101474
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101475
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101476
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101479
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101741
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101743
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101915
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101916
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/102207
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102240
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102307
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102330
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102331
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102484
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102487
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102570
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102571
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102684
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102687
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102965
PRIOR FILING DATE: 1998-10-02
PRIOR APPLICATION NUMBER: 60/103258
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103314
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103315
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103328
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103395
PRIOR FILING DATE: 1998-10-07

/ PRIOR APPLICATION NUMBER: 60/103396
/ PRIOR FILING DATE: 1998-10-07
/ PRIOR APPLICATION NUMBER: 60/103401
/ PRIOR FILING DATE: 1998-10-07
/ PRIOR APPLICATION NUMBER: 60/103449
/ PRIOR FILING DATE: 1998-10-06
/ PRIOR APPLICATION NUMBER: 60/103633
/ PRIOR FILING DATE: 1998-10-08
/ PRIOR APPLICATION NUMBER: 60/103678
/ PRIOR FILING DATE: 1998-10-08
/ PRIOR APPLICATION NUMBER: 60/103679
/ PRIOR FILING DATE: 1998-10-08
/ PRIOR APPLICATION NUMBER: 60/103711
/ PRIOR FILING DATE: 1998-10-08
/ PRIOR APPLICATION NUMBER: 60/104257
/ PRIOR FILING DATE: 1998-10-14
/ PRIOR APPLICATION NUMBER: 60/104987
/ PRIOR FILING DATE: 1998-10-20
/ PRIOR APPLICATION NUMBER: 60/105000
/ PRIOR FILING DATE: 1998-10-20
/ PRIOR APPLICATION NUMBER: 60/105002
/ PRIOR FILING DATE: 1998-10-20
/ PRIOR APPLICATION NUMBER: 60/105104
/ PRIOR FILING DATE: 1998-10-21
/ PRIOR APPLICATION NUMBER: 60/105169
/ PRIOR FILING DATE: 1998-10-22
/ PRIOR APPLICATION NUMBER: 60/105266
/ PRIOR FILING DATE: 1998-10-22
/ PRIOR APPLICATION NUMBER: 60/105693
/ PRIOR FILING DATE: 1998-10-26
/ PRIOR APPLICATION NUMBER: 60/105694
/ PRIOR FILING DATE: 1998-10-26
/ PRIOR APPLICATION NUMBER: 60/105807

Alignment Scores:

Pred. No.: 1.6e-152 Length: 2063
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 11 Gaps: 0

US-09-607-745-9 (1-292) x US-09-946-374-274 (1-2063)

Qy	51	LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln	70
Db	804	CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGTCCAGATCCAG	863
Qy	71	TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla	90
Db	864	TACGACAAACAGCACGCTGTGTGGAGGAGCATCTGGACCCCTGGCTGGTCTCAGGCA	923
Qy	91	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	110
Db	924	GCCCACTGCTTCAGGAAACATACCGATGTGTCAACTGGAAGGTGCGGCGAGGCTCAGAC	983
Qy	111	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro	130
Db	984	AAACTGGGAGCTTCCCATCCCTGGTGTGGCCAAAGATCATCATCATTAATTCAACCCC	1043
Qy	131	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	150
Db	1044	ATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTCCCACTCACTTCTCA	1103
Qy	151	GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	170
Db	1104	GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCC	1163
Qy	171	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	190
Db	1164	CTCTGGATCATTTGATGGGGCTTTTACGAAGCAGAATGGAGGGAAGATGTCTGACATACTG	1223
Qy	191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	210

1224	CTCAGGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCAATGCAGACGATCGGTACCAG	1283
211	GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	230
1284	GGGAAGTCAACCGAGAAGATGATGTGTGCAGGGCATCCGGAAAGGGGGTGTGGACACCTGC	1343
231	GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	250
1344	CAGGTTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGCATC	1403
251	ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer	270
1404	GTTAGCTGGGGCTATGGTTCGGGGGCCCCGAGCACCCAGGAGTATACCAAGGTCATCA	1463
271	AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	284
1464	GCCTATCTCAACTGGATCTCAATGTCTGGAGGCTGAGCTG	1505

RESULT 4

5-10-015-387A-274
Sequence 274, Application US/10015387A
Publication No. US20030135034A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan I.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic Acids Encoding the Same

FILE REFERENCE: P2830P1C54
CURRENT APPLICATION NUMBER: US/10/015,387A

CURRENT FILING DATE: 2001-12-12
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 477

SEO ID NO 274

LENGTH: 2063

TYPE: DNA

ORGANISM: Homo sapiens

ORGANISM: HOMO
b-10-015-387A-274

ignment Scores:

Length:	1.6e-152	2063
Matches:	1265.00	228
Conservative:	98.29%	2
Mismatches:	97.44%	4
Indels:	99.12%	0
Gaps:	12	0

-09-607-745-9 (1-292) x US-10-015-387A-274 (1-2063)

51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70

804 CGTGTGCTGGTGGGAGGAGGCCTCTGTGGATTCTTGGCCCTTGGCAGGTTCAGCATCCAG 863

71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90

864 TACGACAAACAGCAGCTCTGTGGAGGAGCATCTCGACCCCACTGGGTCTTCACGGCA 923

91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110

924 GCCCACTGCTTCAGGAACATACCGATGTGTCAACTGGAGTGCGGCAGGCTCAGAC 983

QY	111	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro	130
Db	984	AAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATTTGAATTCACCCC	1043
QY	131	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	150
Db	1044	ATGTACCCCAAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCACACTCACTTTCTCA	1103
QY	151	GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	170
Db	1104	GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA	1163
QY	171	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	190
Db	1164	CTCTGGATCAITGGATGGGGCTTTACGAAGCAGAATGGAGGAAGATGTCTGCACATCTG	1223
QY	191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	210
Db	1224	CTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCAAATGCAGACGATGCGTACCAG	1283
QY	211	GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	230
Db	1284	GGGGAAGTCACCGAGAAGATGATGTGTGAGGCATCCCGGAAGGGGTGTGCACACCTGC	1343
QY	231	GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	250
Db	1344	CAGGGTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGCATC	1403
QY	251	ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer	270
Db	1404	GTTAGCTGGGGCTATGGCTCGGGGGGCCGAGCACCCAGGAGTATACACCAAGGTCTCA	1463
QY	271	AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	284
Db	1464	GCCTATCTCAACTGGATCTACAAATGCTCTGGAAGGCTGAGCTG	1505

RESULT 5

US-10-063-735-177

: Sequence 111 Application 111/1005372E

Publication No. US20030138882A1

; PUBLICATION NO: US20
; GENERAL INFORMATION:

GENERAL INFORMATION:

APPLICANT: Eaton, Dan L.

APPLICANT: Filvaroff, Ellen

APPLICANT: Gerritsen, Mary E.

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

1. TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACID POLYMERIZATION

1. TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE
2. TITLE OF INVENTION: ACIDS ENCODING THE SAME

FILE OF INVENTION: ACIDS ENCODING THE S
; FILE REFERENCE: P3230R1C7

FILE REFERENCE: F3230K1C1
: CURRENT APPLICATION NUMBER: IIS/1

CURRENT AFFILIATION NUMBER: 08/10/063, 133
: CURRENT FILING DATE: 2003-05-08

CURRENT FILING DATE: 200

DATE	PILOT AB.	NUMBER OF CE
10/1/50	100	100
10/2/50	100	100
10/3/50	100	100
10/4/50	100	100
10/5/50	100	100
10/6/50	100	100
10/7/50	100	100
10/8/50	100	100
10/9/50	100	100
10/10/50	100	100
10/11/50	100	100
10/12/50	100	100
10/13/50	100	100
10/14/50	100	100
10/15/50	100	100
10/16/50	100	100
10/17/50	100	100
10/18/50	100	100
10/19/50	100	100
10/20/50	100	100
10/21/50	100	100
10/22/50	100	100
10/23/50	100	100
10/24/50	100	100
10/25/50	100	100
10/26/50	100	100
10/27/50	100	100
10/28/50	100	100
10/29/50	100	100
10/30/50	100	100
10/31/50	100	100

; NUMBER OF SEQ
; SEQ TO NO 111

SEQ ID NO 1

; LENGTH: 2063
TIME 0000

TYPE: DNA

Alignment Scores:

Pred. No.:	1.6e-152	Length:	2063
Score:	1266.00	Matches:	228
Percent Similarity:	98.29%	Conservative:	2
Best Local Similarity:	97.44%	Mismatches:	4
Query Match:	79.12%	Indels:	0
DB:	12	Gaps:	0

US-09-607-745-9 (1-292) x US-10-063-735-111 (1-2063)

```
51  LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
    ::::|||||
804  CGTGTGGTGGTGGGAGGAGGCCTCTGTGGATTCTTGGCCCTTGGCAGGTTCAGCATCCAG 863

71  TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
    |||||
864  TACGACAAACAGACAGCTGTGTGGAGGAGCATCCTGGACCCCACTGGGTCTCTCACGGCA 923

91  AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
    |||||
924  GCCCACTGCTTCAGAAACATACCGATGTGTCAACTGGAAGTGGCGGCGAGGCTCAGAC 983

111  LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
    |||||
984  AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCTGGCCAAAGATCATCATATTGAATTCACCC 1043

131  MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
    |||||
1044  ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCTCCACTCTCTCTCA 1103

151  GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
    |||||
1104  GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163

171  LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
    |||||
1164  CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACTG 1223

191  LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
    |||||
1224  CTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGTCAATGCAGACGATGCGGTACCAG 1283

211  GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
    |||||
1284  GGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGGAAGGGGTGTGGACACCTGC 1343

231  GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
    |||||
1344  CAGGGTGACACTGGTGGGCCCTGTATACCAATCTGACCAAGTGGCATGTGGTGGGCATC 1403

251  ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
    |||||
1404  GTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCCGAGGAGTATACCAAGGTCTCA 1463

271  AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
    |||||
1464  GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505
```

RESULT 6

5-10-006-130A-274
Sequence 274, Application US/10006130A
Publication No. US20030148375A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan I.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2830P1C7
CURRENT APPLICATION NUMBER: US/10/006,130A
CURRENT FILING DATE: 2002-03-19
Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 274
; LENGTH: 2063
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-006-130A-274

Alignment Scores:

Pred. No.: 1.6e-152 Length: 2063
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 12 Gaps: 0

US-09-607-745-9 (1-292) x US-10-006-130A-274 (1-2063)

```
QY 51  LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
    ::::|||||
Db 804  CGTGTGGTGGTGGGAGGAGGCCTCTGTGGATTCTTGGCCCTTGGCAGGTTCAGCATCCAG 863

QY 71  TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
    |||||
Db 864  TACGACAAACAGACAGCTGTGTGGAGGAGCATCCTGGACCCCACTGGGTCTCTCACGGCA 923

QY 91  AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
    |||||
Db 924  GCCCACTGCTTCAGAAACATACCGATGTGTCAACTGGAAGGTGCGGGCAGGCTCAGAC 983

QY 111  LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
    |||||
Db 984  AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATATTGAATTCACCC 1043

QY 131  MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
    |||||
Db 1044  ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCTCCACTCTCTCTCA 1103

QY 151  GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
    |||||
Db 1104  GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163

QY 171  LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
    |||||
Db 1164  CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACTG 1223

QY 191  LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
    |||||
Db 1224  CTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGTCAATGCAGACGATGCGGTACCAG 1283

QY 211  GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
    |||||
Db 1284  GGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGGAAGGGGTGTGGACACCTGC 1343

QY 231  GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
    |||||
Db 1344  CAGGGTGACAGTGTGGGCCCTGTATACCAATCTGACCAAGTGGCATGTGGTGGGCATC 1403

QY 251  ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
    |||||
Db 1404  GTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCCGAGGAGTATACCAAGGTCTCA 1463

QY 271  AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
    |||||
Db 1464  GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505
```

RESULT 7

US-10-199-672-329
; Sequence 329, Application US/10199672
; Publication No. US20030148442A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc

;	PRIOR APPLICATION NUMBER:	60/101743
;	PRIOR FILING DATE:	1998-09-24
;	PRIOR APPLICATION NUMBER:	60/101743
;	PRIOR FILING DATE:	1998-09-24
;	PRIOR APPLICATION NUMBER:	60/101915
;	PRIOR FILING DATE:	1998-09-24
;	PRIOR APPLICATION NUMBER:	60/101916
;	PRIOR FILING DATE:	1998-09-24
;	PRIOR APPLICATION NUMBER:	60/102207
;	PRIOR FILING DATE:	1998-09-29
;	PRIOR APPLICATION NUMBER:	60/102240
;	PRIOR FILING DATE:	1998-09-29
;	PRIOR APPLICATION NUMBER:	60/102307
;	PRIOR FILING DATE:	1998-09-29
;	PRIOR APPLICATION NUMBER:	60/102330
;	PRIOR FILING DATE:	1998-09-29
;	PRIOR APPLICATION NUMBER:	60/102331
;	PRIOR FILING DATE:	1998-09-29
;	PRIOR APPLICATION NUMBER:	60/102484
;	PRIOR FILING DATE:	1998-09-30
;	PRIOR APPLICATION NUMBER:	60/102487
;	PRIOR FILING DATE:	1998-09-30
;	PRIOR APPLICATION NUMBER:	60/102570
;	PRIOR FILING DATE:	1998-09-30
;	PRIOR APPLICATION NUMBER:	60/102571
;	PRIOR FILING DATE:	1998-09-30
;	PRIOR APPLICATION NUMBER:	60/102684
;	PRIOR FILING DATE:	1998-10-01
;	PRIOR APPLICATION NUMBER:	60/102687
;	PRIOR FILING DATE:	1998-10-01
;	PRIOR APPLICATION NUMBER:	60/102965
;	PRIOR FILING DATE:	1998-10-02
;	PRIOR APPLICATION NUMBER:	60/103258
;	PRIOR FILING DATE:	1998-10-06
;	PRIOR APPLICATION NUMBER:	60/103314
;	PRIOR FILING DATE:	1998-10-07
;	PRIOR APPLICATION NUMBER:	60/103315
;	PRIOR FILING DATE:	1998-10-07
;	PRIOR APPLICATION NUMBER:	60/103328
;	PRIOR FILING DATE:	1998-10-07
;	PRIOR APPLICATION NUMBER:	60/103395
;	PRIOR FILING DATE:	1998-10-07
;	PRIOR APPLICATION NUMBER:	60/103396
;	PRIOR FILING DATE:	1998-10-07
;	PRIOR APPLICATION NUMBER:	60/103401
;	PRIOR FILING DATE:	1998-10-07
;	PRIOR APPLICATION NUMBER:	60/103449
;	PRIOR FILING DATE:	1998-10-06
;	PRIOR APPLICATION NUMBER:	60/103633
;	PRIOR FILING DATE:	1998-10-08
;	PRIOR APPLICATION NUMBER:	60/103678
;	PRIOR FILING DATE:	1998-10-08
;	PRIOR APPLICATION NUMBER:	60/103679
;	PRIOR FILING DATE:	1998-10-08
;	PRIOR APPLICATION NUMBER:	60/103711
;	PRIOR FILING DATE:	1998-10-08
;	PRIOR APPLICATION NUMBER:	60/104257
;	PRIOR FILING DATE:	1998-10-14
;	PRIOR APPLICATION NUMBER:	60/104987
;	PRIOR FILING DATE:	1998-10-20
;	PRIOR APPLICATION NUMBER:	60/105000
;	PRIOR FILING DATE:	1998-10-20
;	PRIOR APPLICATION NUMBER:	60/105002
;	PRIOR FILING DATE:	1998-10-20
;	PRIOR APPLICATION NUMBER:	60/105104
;	PRIOR FILING DATE:	1998-10-21
;	PRIOR APPLICATION NUMBER:	60/105169
;	PRIOR FILING DATE:	1998-10-22
;	PRIOR APPLICATION NUMBER:	60/105266
;	PRIOR FILING DATE:	1998-10-22
;	PRIOR APPLICATION NUMBER:	60/105693
;	PRIOR FILING DATE:	1998-10-26
;	PRIOR APPLICATION NUMBER:	60/105694

151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTTTGTATGAGGAGCTCACTCCAGCCACCCCA 1163
171 LeuTIPileIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
1164 CTCTGGATCATGGATGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATACTG 1223
191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
1224 CTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCAATGACAGACGATGCGTACCAG 1283
211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
1284 GGGGAAGTCACCGAGAAGATGATGTGTCAGGCATCCCGGAAGGGGTGTGGACACCTGC 1343
231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
1344 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGCATC 1403
251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
1404 GTTAGCTGGGCTATGGCTCGGGGGCCGAGCACCCCGAGGAGTATACACCAAGGTCTCA 1463
271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
1464 GCCTATCTCACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505

RESULT 10

S-10-194-457-329
Sequence 329, Application US/10194457
Publication No. US20030153037A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C296
CURRENT APPLICATION NUMBER: US/10/194,457
CURRENT FILING DATE: 2002-07-11
PRIOR APPLICATION NUMBER: 10/052586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 612

LENGTH: 2063
TYPE: DNA
ORGANISM: Homo Sapien
US-10-194-457-329
Alignment Scores:
Pred. No.: 1.6e-152 Length: 2063
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 12 Gaps: 0
US-09-607-745-9 (1-292) x US-10-194-457-329 (1-2063)
QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
Db 804 CGTGTGGTGGTGGGAGGAGGCCTCTGTGGATTCTTGGCCTTGGCAGGTCCAGATCCAG 863
QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
Db 864 TACGACAAACAGCACGCTCTGTGGAGGGAGCATCTCTGACCCCCCACTGGGTCTCAGCGCA 923
QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
Db 924 GCCCACTGCTTCAGGAACATACATCGATGTGTCAACTGGAAGGTGCGGCGAGGCTCAGAC 983
QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
Db 984 AAACCTGGGCAGCTTCCCATCCCTGGTGGTGGCCCAAGATCATCATATTGAATTCAACCCC 1043
QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
Db 1044 ATGTACCCCAAGACATGACATCGCCCTCATGAGAGTGCAGTTCCCACTCACTTTCTCA 1103
QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
Db 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163
QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
Db 1164 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATACTG 1223
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
Db 1224 CTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCAATGACAGCGATGCGTACCAG 1283
QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
Db 1284 GGGGAAGTCACCGAGAAGATGATGTGTGAGGATCCCGGAAGGGGTGTGGACACCTGC 1343
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
Db 1344 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGCATC 1403
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
Db 1404 GTTAGCTGGGCTATGGCTCGGGGGCCGAGCACCCCGAGGAGTATACACCAAGGTCTCA 1463
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1464 GCCTATCTCACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505

RESULT 11

US-10-184-642-329
Sequence 329, Application US/10184642
Publication No. US20030157635A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.

APPLICANT: Gurney,Austin L.
APPLICANT: Pan,James
APPLICANT: Smith,Victoria
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C194
CURRENT APPLICATION NUMBER: US/10/184,642
CURRENT FILING DATE: 2002-06-27
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 329
LENGTH: 2063
TYPE: DNA
ORGANISM: Homo Sapien
S-10-184-642-329

Alignment Scores:
Pred. No.: 1.6e-152 Length: 2063
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
B: 12 Gaps: 0
S-09-607-745-9 (1-292) x US-10-184-642-329 (1-2063)

51 LysileValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
: : : : :
804 CGTGTGGTGGTGGGAGGAGGCGCTCTGTGGATTCTTGGCCTTGGCAGGTCAGCATCCAG 863
71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
: : : : :
864 TACGACAAACAGCAGCTCTGTGGAGGAGCATCCTGGACCCCTGGTCTCTACGGCA 923
91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
: : : : :
924 GCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAGGTGGGGCAGGCTCAGAC 983
111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
: : : : :
984 AAACCTGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATATTGAATTCACCC 1043
131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
: : : : :
1044 ATGTACCCAAAGACAATGACATCGCCTCATGAAGCTGCATGCCACTCATTCTCTCA 1103
151 GlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAlaThrPro 170
: : : : :
1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCC 1163
171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
: : : : :
1164 CTCTGGATCATTGGATGGGCTTTACGAAGCAGAAATGGAGGGAAGATGCTGACATACTG 1223
191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
: : : : :
1224 CTGCAGGCTCAGTCCAGGTCATTGACAGCACACGGTGCAATGCAGACGATGGGTACCAG 1283
211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
: : : : :
1284 GGGGAAGTCAACGAGAAGATGATGTGTGCAGGCATCCCGAAGGGGTGTGGACACCTGC 1343
231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
: : : : :
1344 CAGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGCATC 1403
251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
: : : : :
1404 GTTAGCTGGGCTATGGCTGGGGGGGGGGAGGACCCAGGAGTATACACCAAGGTCTCA 1463

QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
: : : : :
Db 1464 GCCTATCTCAACTGGATCTACAATGTCTGAAGGCTGAGCTG 1505
RESULT 12
US-10-196-747-329
; Sequence 329, Application US/10196747
; Publication No. US20030162250A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C346
; CURRENT APPLICATION NUMBER: US/10/196,747
; CURRENT FILING DATE: 2002-07-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 329
; LENGTH: 2063
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-196-747-329

Alignment Scores:
Pred. No.: 1.6e-152 Length: 2063
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 12 Gaps: 0
US-09-607-745-9 (1-292) x US-10-196-747-329 (1-2063)

QY 51 LysileValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
: : : : :
Db 804 CGTGTGGTGGTGGGAGGAGGCGCTCTGTGGATTCTTGGCCTTGGCAGGTCAGCATCCAG 863
71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
: : : : :
Db 864 TACGACAAACAGCAGCTCTGTGGAGGAGCATCCTGGACCCCTGGTCTCTACGGCA 923
91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
: : : : :
Db 924 GCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAGGTGGGGCAGGCTCAGAC 983
111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
: : : : :
Db 984 AAACCTGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATATTGAATTCACCC 1043
131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
: : : : :
Db 1044 ATGTACCCAAAGACAATGACATCGCCTCATGAAGCTGCAGTTCCTTCTCTCA 1103
151 GlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAlaThrPro 170
: : : : :
Db 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCC 1163
171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
: : : : :
Db 1164 CTCTGGATCATTGGATGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATACTG 1223
191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
: : : : :

b 1224 CTCAGGGCGTCAGTCCAGGTCATTGACAGCACACCGTGTCAATGCAGACCATGCGTACCAG 1283
y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
b 1284 GGGGAAGTCCCGAGAAGATGATGTGTCAGGCATCCCGAAGGGGTGTGGACACCTGC 1343
y 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
b 1344 CAGGGTGACAGTGGTGGCCCTGATGTACCAATGTGACCAAGTGGCATGTGGGCATC 1403
y 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
b 1404 GTTAGCTGGGCTATGGCTGGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1463
y 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
b 1464 GCCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1505

RESULT 13

S-10-015-392A-274
Sequence 274, Application US/10015392A
Publication No. US20030166901A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan I.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2830P1C58
CURRENT APPLICATION NUMBER: US/10/015,392A
CURRENT FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: 60/098716
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098723
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098749
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098750
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098803
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098821
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098843
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/099536
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 477
SEQ ID NO 274
LENGTH: 2063
TYPE: DNA
ORGANISM: Homo sapiens
S-10-015-392A-274

Lignment Scores: 1.6e-152 Length: 2063
red. No.: 1266.00 Matches: 228
core:

Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 12 Gaps: 0
US-09-607-745-9 (1-292) x US-10-015-392A-274 (1-2063)
QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
Db 804 CGTGTGGTGGTGGGAGGAGGCCCTCTGTGGATTCTTGGCCCTTGGCAGGTTCAGCATCCAG 863
QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
Db 864 TACGACAAACAGACAGCTCTGTGGAGGAGCATCCTGGACCCCCACTGGGTCTCTCAGCGCA 923
QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
Db 924 GCCCACTGCTCAGGAACAATACCCGATGTGTTCACCTGGAGAGTGGGGCAGGCTCAGAC 983
QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
Db 984 AAAGTGGGAGCTTCCCCTCCCTGGCTGTGGCCCAAGATCATCATCATCATCAACCC 1043
QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
Db 1044 ATGTACCCCAAGACAATGACATCCGCTCATGAAGTGCAGTCCCACTCACTTTCTCA 1103
QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
Db 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGTCACTCCAGCCACCCCA 1163
QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
Db 1164 CTCTGGATCATGGATGGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATCTG 1223
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
Db 1224 CTGCAGGCGTCAGTCCAGGTCAATGACAGCACACGGTGCATATGCAGACGATGCGTACCAG 1283
QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
Db 1284 GGGGAAGTCCCGAGAAGATGATGTGTGTCAGGCATCCCGAAGGGGGTGTGGACACCTGC 1343
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
Db 1344 CAGGGTGACAGTGGTGGGCCCTCTGATACCAATCTGACCAATCTGACCAAGTGGCATGTGGTGGGCATC 1403
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
Db 1404 GTTAGCTGGGCTATGGCTGGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1463
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1464 GCCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1505

RESULT 14

US-10-017-253A-274
; Sequence 274, Application US/10017253A
; Publication No. US20030166055A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2830PIC62
CURRENT APPLICATION NUMBER: US/10/017,253A
CURRENT FILING DATE: 2001-12-13
PRIOR APPLICATION NUMBER: 60/098716
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098723
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098749
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098750
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098803
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098821
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098843
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/099536
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 477
SEQ ID NO 274
LENGTH: 2063
TYPE: DNA
ORGANISM: Homo sapiens
S-10-017-253A-274

Alignment Scores:
red. No.: Length: 1.6e-152 2063
core: Matches: 1266.00 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
Gaps: 0
3:
S-09-607-745-9 (1-292) x US-10-017-253A-274 (1-2063)
Y 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
: : : : :
b 804 CGTGTGGTGGTGGGAGGAGGCCTCTGTGGATTCTTGGCCCTTGGCAGGTCAGCATCCAG 863
Y 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
: : : : :
b 864 TACGACAAACAGCACGCTGTGGAGGAGCATCCTGGACCCCACTGGGTCTCTCAGGCA 923
Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
: : : : :
b 924 GCCCACTGCTTCAGGAACATACCCGATGTGTTCAACTGGAAGTGGCGGCGAGGCTCAGAC 983
Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
: : : : :
b 984 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATATTGAATTCACACCC 1043
Y 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
: : : : :
b 1044 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGTCGAGTCCCACTCACTTCTCA 1103
Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
: : : : :
b 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGTCACCTCAGCCACCCCA 1163
Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
: : : : :
b 1164 CTCTGGATCATGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATACTG 1223
Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210

Db 1224 CTGCAGCGCTCAGTCCAGGTCATTGACAGCACACGGTGCATGCAATGACAGCATCGTACCAG 1283
QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThrCys 230
: : : : :
Db 1284 GGGGAAGTCACCGAAGATGATGTGTGAGGCATCCCGGAAGGGGGTGTGGACACCTGC 1343
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
: : : : :
Db 1344 CAGGGTGACAGTGGTGGCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGCATC 1403
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
: : : : :
Db 1404 GTTAGCTGGGGCTATGGCTGCGGGGGCCGAGCACCCCGAGGATATACACCAAGGTCTCA 1463
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
: : : : :
Db 1464 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505
RESULT 15
US-10-173-689-329
; Sequence 329, Application US/10173689
; Publication No. US20030166104A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C10
; CURRENT APPLICATION NUMBER: US/10/173,689
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 329
; LENGTH: 2063
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-173-689-329

Alignment Scores:
Pred. No.: Length: 1.6e-152 2063
Score: Matches: 1266.00 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: Gaps: 0
US-09-607-745-9 (1-292) x US-10-173-689-329 (1-2063)
QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
: : : : :
Db 804 CGTGTGGTGGTGGGAGGAGGCCTCTGTGGATTCTTGGCCCTTGGCAGGTCAGCATCCAG 863
QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
: : : : :
Db 864 TACGACAAACAGCACGCTGTGTGGAGGAGCATCCTGGACCCCACTGGGTCTCTCAGGCA 923
QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
: : : : :
Db 924 GCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGTGGCGGCGAGGCTCAGAC 983
QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
: : : : :
Db 984 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATATTGAATTCACACCC 1043

Y 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
b 1044 ATGTACCCCAAGACATGACATGCCCTCATGAAGCTGCAGTCCCACTTCTCA 1103
Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
b 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGAGCTCACTCCAGCCACCCCA 1163
Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
b 1164 CTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAAATGGAGGGAAGATGCTGTGACATACTG 1223
Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
b 1224 CTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGGTGCAATGCAGAGCATGCGTACCAG 1283
Y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
b 1284 GGGGAAGTCACCGAGAAGATGATGTGTGAGGCATCCCGGAAGGGGTGTGGACACCTGC 1343
Y 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
b 1344 CAGGGTGACAGTGGTGGSCCCTGATGTATCCATCTGACCAAGTGGCATGTGGTGGGCATC 1403
Y 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
b 1404 GTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCCGAGGAGTATACCAAGGTCTCA 1463
Y 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
b 1464 GCCTATCTCAACTGGATCTACATGTCTTGGAGGCTGAGCTG 1505

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

protein - nucleic search, using frame_plus_p2n model

run on: November 23, 2003, 07:53:00 ; Search time 2075.73 Seconds
(without alignments)
3418.993 Million cell updates/sec

title: US-09-607-745-9
perfect score: 1600
sequence: 1 MDSKSSQKSRLLLLLVSN.....LNWIYVWKAELSRHHHHH 292

scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

searched: 22781392 seqs, 12152238056 residues
total number of hits satisfying chosen parameters: 45562784

minimum DB seq length: 0
maximum DB seq length: 2000000000
post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

command line parameters:
MODEL=frame+ p2n.model -DEV=xlh
Q2=/cgn2_1/USPTO.spool/US09607745/runat_21112003_144346_22229/app_query.fasta_1.1038
DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
LOCALALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
JSER=US09607745@cgn_1_3596@runat_21112003_144346_22229 -NCPU=6 -ICPU=3
NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1102.5	68.9	975	10	BG288427
2	1050	65.6	2177	11	AK078890 Mus muscu
3	1021	63.8	895	13	BU522841 AGENCOURT
4	968	60.5	1973	11	BC004855 Homo sapi
5	895	55.9	689	12	BG966811 602834306
6	885	55.3	515	14	CB142902 K-EST0196
7	840	52.5	569	12	BM686292 UI-E-CR0-
c	816	51.0	659	9	AI924527 wn61c07.x
c	808	50.5	648	9	AI924182 wn53f11.x
10	797	49.8	563	12	BM773306 K-EST0057
11	797	49.8	572	12	BM746729 K-EST0021
12	797	49.8	578	12	BM747250 K-EST0021
13	797	49.8	593	12	BM788452 K-EST0067
14	797	49.8	596	12	BM772743 K-EST0056
15	797	49.8	614	12	BM788163 K-EST0067
16	797	49.8	633	12	BM764659 K-EST0046
17	797	49.8	691	12	BM763697 K-EST0045
18	797	49.8	692	12	BM772754 K-EST0056
19	797	49.8	700	12	BM772620 K-EST0056
20	783	48.9	548	12	BM746721 K-EST0021
21	780	48.8	920	13	BQ922656 AGENCOURT
22	687	42.9	408	12	BM762292 K-EST0043
c	657	41.1	584	9	AI597567 tr92a03.x
24	656.5	41.0	607	9	AA143746 zo67e02.r
25	655	40.9	711	10	BE937482 RC4-ST027
26	652	40.8	435	10	BF811390 CM2-CI017
27	648.5	40.5	644	12	BM791640 K-EST0071
28	648	40.5	481	9	AA143761 zo67g02.r
29	637	39.8	721	13	BU600097 AGENCOURT
30	618	38.6	425	10	AW908659 uf81c03.y
31	609	38.1	956	13	BU523218 AGENCOURT
32	603	37.7	542	10	AW913122 uf49b12.y
33	589	36.8	569	9	AW751450 MRI-CT005
34	558	34.9	702	12	BM695610 UI-E-CQ1-
35	551	34.4	498	9	AI158537 ud25e02.r
c	538.5	33.7	1835	11	BC048135 Danio rer
37	531.5	33.2	1172	14	CD508404 CDA90-F11
38	519	32.4	2895	11	AK029714 Mus muscu
39	514.5	32.2	1173	14	CD508765 CDA92-F09
40	506	31.6	1749	11	BC035623 Homo sapi
c	502	31.4	586	9	AW613143 hh42c04.x
42	500	31.2	683	13	BU290098 604163681
43	493	30.8	2055	11	BC028065 Homo sapi
44	491	30.7	3035	11	AK004939 Mus muscu
45	481	30.1	816	13	BU397449 603536894

ALIGNMENTS

RESULT 1
BG288427
LOCUS
DEFINITION
602388091F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4516990 5',
975 bp mRNA linear EST 21-FEB-2001
mRNA sequence.
ACCESSION
BG288427
VERSION
BG288427.1 GI:13043459
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 975)


```
Y 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
b 906 TACGACAAACAGCAGCTCTGTGGAGGAGCATCCTGGACCCCACTGGGTCTCAGCGCA 965
Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
b 966 GCCCACTGCTTCAGGAAACATACCGATGTGTCAACTGGAAGGTGGGGCAGGCTCAGAC 1025
Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
b 1026 AAACCTGGGCAGCTCCCATCCCTGGCTGTGGCCCAAGATCATCATCATTAATTCAACCCC 1085
Y 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
b 1086 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCTCACTTCTCA 1145
Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
b 1146 GGCACAGTACGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1205
Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
b 1206 CTCTGGATCATGTGATGGGCTTTACGAAGCAGAAATGGA----- 1244
Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
b 1244 ----- 1244
Y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
b 1244 ----- 1244
Y 231 GlnGly-AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
b 1245 ---GGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCACTGGCATGTGGTGGCAT 1301
Y 250 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe 270
b 1302 CGTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTC 1361
Y 270 rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
b 1362 AGCCTATCTCAACTGGATCTACAAATCTGTGGAAGGCTGAGCTG 1404
```

```
RESULT 5
3966811
OCUS
DEFINITION
  BG966811
  602834306F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4989048 5',
  mRNA sequence.
  BG966811
  BG966811.1 GI:14354448
  EST.
  Mus musculus (house mouse)
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 689)
  NIH-MGC http://mgi.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished
  Contact: Robert Strausberg, Ph.D.
  Email: cgabs-remail.nih.gov
  Tissue Procurement: Jeffrey E. Green, M.D.
  cDNA Library Preparation: Life Technologies, Inc.
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLAM11002 row: k column: 01
  High quality sequence stop: 680.
  Location/Qualifiers
```

```
source
1. .689
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4989048"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Co24"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 160 a 189 c 188 g 152 t
ORIGIN

Alignment Scores:
Pred. No.: 3.78e-91 Length: 689
Score: 895.00 Matches: 167
Percent Similarity: 88.43% Conservative: 24
Best Local Similarity: 77.31% Mismatches: 25
Query Match: 55.94% Indels: 2
DB: 12 Gaps: 0

US-09-607-745-9 (1-292) x BG966811 (1-689)

QY 75 HisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAlaHisCysPhe 94
Db 2 CATGTCTGTGGTGGAGCATCTCTGGATCCCCACTGGATCTCTCAGCAGCCACTGCTTC 61
QY 95 ArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAspLysLeuGlySer 114
Db 62 AGGAAGTATCTTGTATGTCTAAGCTGGAAGGTGAGGGCAGGCTCAAACATACTGGGTAA 121
QY 115 PheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnProMetTyrProLys 134
Db 122 TCTCCATCTCTGCTGTGGCCAAAGATCTTCATCGTGAACCAATCTCTGTATCCCCAAA 181
QY 135 AspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArg 154
Db 182 GAGAAGGACATTGCCCTTGTAAAGCTGCAGATGCCACTCACATTCTCAGGCTCAGTCAGG 241
QY 155 ProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrProLeuTrpIleIle 174
Db 242 CCCATCTGCCCTGCCCTTCTCTGATGAGGTGCTTGTCCAGCCACACCATCTCTGGGTCA 301
QY 175 GlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSer 194
Db 302 GGATGGGGCTTTACAGAAAGAAACGGAGGAAAGATGTCTGACATGTCTACTGCAGGCATCA 361
QY 195 ValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGlnGlyGluValThr 214
Db 362 GTCCAGGTCAATCGACAGCACAGGTGCAATGCAAGGATGCCTACGAAGGGAAGTGACC 421
QY 215 GluLysMetMetCysAlaGlyIleProGluGlyValAspThrCysGlnGlyAspSer 234
Db 422 GCTGAGATGCTGTGTGAGGTACCCCAACAGGGTGGCAAGGACACCTGCCA-GGTGACAGT 480
QY 235 GlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIleValSerTrpGly 254
Db 481 GGTGGGCTTTGATGTACCATCTCTGACAAGTGGCAGGTAGTAGGCATCTGTGAGCTGGGG 540
QY 255 TyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsn 274
Db 541 CATGGATCGGGCGG-CCAAGTACTCTGGAGTGTATACCAAGGTCACTGCCTATCTCAAC 599
QY 275 TrpIleTyrAsnValTrpLysAlaGluLeuSerArgHisHisHis 290
Db 600 TGGATCTACAATGTTCCGGAAGTCTGAGATGTAACTGACGCTGCCGTCCCCAC 647
```

```
RESULT 6
CB142902
LOCUS
DEFINITION K-EST0196799 L11SNU354s1 Homo sapiens cDNA clone L11SNU354s1-20-D07
```



```
5', mRNA sequence.
CESSION      CB142902
VERSION      CB142902.1  GI:28119586
EYWORDS      EST.
SOURCE       Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 515)
AUTHORS      Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
              Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
              Kim,Y.S.
TITLE        21C Frontier Korean EST Project 2001
JOURNAL      Unpublished
COMMENT      Contact: Kim Ys
              Genome Research Center
              Korea Research Institute of Bioscience & Biotechnology
              52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
              Tel: +82-42-860-4470
              Fax: +82-42-860-4409
              Email: yongsung@mail.kribb.re.kr
              Plate: 20 row: D column: 07
              High quality sequence stop: 515.
              Location/Qualifiers
                1..515
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="L11SNU354s1-20-D07"
                  /sex="M"
                  /tissue_type="Liver"
                  /cell_line="Polygonal"
                  /lab_host="Top10F"
                  /clone_lib="L11SNU354s1"
                  /note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI;
                  Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
                  bacterial alkaline phosphatase (BAP) and then decapped
                  with tobacco acid pyrophosphatase (TAP). The decapped
                  intact mRNA was ligated with DNA-RNA linker including
                  EcoRI site by treatment of T4 RNA ligase and the first
                  strand cDNA was synthesized from oligo dT-selected mRNA by
                  priming with dT-tailed vector. The dT-tailed vector was
                  adjusted to have about 60nt. The cDNA vector was
                  circularized with E. coli DNA ligase after digestion of
                  EcoRI which site is also included in vector. An RNA strand
                  converted to a DNA strand by Okayama-Berg method. The
                  obtained cDNA vectors were used for transformation of
                  competent cells E. coli Top10F, by electroporation method.
                  The cDNA libraries constructed by this method are
                  full-length enriched cDNA library. After analyzing and
                  sequencing about 2,000 - 3,000 colonies in original cDNA
                  library, the abundant cDNAs were selected and amplified by
                  PCR reaction using vector region primer including T7
                  promoter as 5' primer and N(dT)14 as 3' primer. The PCR
                  products were used as template for synthesis of
                  biotinylated single stranded RNA by in vitro transcription
                  reaction. The synthesized RNA probes were hybridized with
                  antisense single stranded cDNAs prepared from original
                  library and incubated with avidin-gel. After removing
                  DNA-RNA hybrids by centrifuge, the subtracted cDNA
                  libraries were constructed by transfection of the
                  remaining DNA into competent cells E. coli Top10F, with
                  electroporation method."
BASE COUNT   116 a 139 c 150 g 110 t
ORIGIN
Alignment Scores:
  ed. No.:      3.31e-90      Length:      515
  core:         885.00      Matches:     160
  Percent Similarity: 100.00%      Conservative: 0
  st Local Similarity: 100.00%      Mismatches:  0
  erty Match:    55.31%      Indels:     0

DB:          14          Gaps:          0
US-09-607-745-9 (1-292) x CB142902 (1-515)
QY          125  I I e I l e G l u P h e A s n P r o M e t T y r P r o L y s A s p A s n A s p I l e A l a L e u M e t L y s L e u G l n 144
Db          2   A T C A T T G A A T T C A A C C C C A T G T A C C C C A A G A C A A T G A C A T G C C C T C A T G A A G C T G C A G 61
QY          145  P h e P r o L e u T h r P h e S e r G l y T h r V a l A r g P r o I l e C y s L e u P r o P h e P h e A s p G l u G l u 164
Db          62  T T C C C A C T C A C T T T C T C A G G C A C A G T C A G G C C C A T C T C T G C C C T T C T T G A T G A G G A G 121
QY          165  L e u T h r P r o A l a T h r P r o L e u T h r P i l e I l e G l y T r p G l y P h e T h r L y s G l n A s n G l y G l y 184
Db          122  C T C A C T C C A G C C A C C C C A C T C T G G A T C A T T G G A T G G G G C T T T A C G A A G C A G A T G G A G G G 181
QY          185  L y s M e t S e r A s p I l e L e u L e u G l n A l a S e r V a l G l n V a l I l e A s p S e r T h r A r g C y s A s n 204
Db          182  A A G A T G T C T G A C A T A C T G C T G C A G G C G T C A G T C C A G G T C A T T G A C A G C A C A C G T G C A A T 241
QY          205  A l a A s p A l a T y r G l n G l u V a l T h r G l u L y s M e t M e t C y s A l a G l y I l e P r o G l u 224
Db          242  G C A G A C G A T C G T A C C A G G G G A A G T C A C C G A G A A G A T G A T G T G T G C A G G C A T C C C G G A A 301
QY          225  G l y G l y V a l A s p T h r C y s G l n G l y A s p S e r G l y G l y P r o L e u M e t T y r G l n S e r A s p G l n 244
Db          302  G G G G T G T G G A C A C C T G C C A G G T G A C A G T G T G G G C C C C T G A T G T A C C A A T C T G A C C A G 361
QY          245  T r p H i s V a l G l y I l e V a l S e r T r p G l y T y r G l y C y s G l y G l y P r o S e r T h r P r o G l y 264
Db          362  T G G A T G T G T G G G C A T C G T A G T T G G G C T A T G G T G G G G C C C G A G C C C A G G A 421
QY          265  V a l T y r T h r L y s V a l S e r A l a T y r L e u A s n T r p I l e T y r A s n V a l T r p L y s A l a G l u L e u 284
Db          422  G T A T A C A C C A A G G T C T C A G C T A T C T C A A C T G G A T C T A C A A T G T C T G G A A G G C T G A G C T G 481

RESULT 7
BM686292
LOCUS      BM686292
DEFINITION UI-E-CRO-adk-g-09-0-UI.r1 UI-E-CRO Homo sapiens cDNA clone
            UI-E-CRO-adk-g-09-0-UI 5', mRNA sequence.
ACCESSION  BM686292
VERSION    BM686292.1  GI:18999550
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 569)
AUTHORS    Ronaldo M.F., Lennon, G. and Soares, M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    97044477
PUBMED     8889548
COMMENT    Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 Reverse.
            Location/Qualifiers
              1..569
                /organism="Homo sapiens"
                /mol_type="mRNA"

FEATURES
            source
```


b 539 GGATGGGGCTTTACGACGAGAATGGAGGGAAGATGCTTGACATACTGCTGCAGGCGTCA 480
Y 195 ValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGlnGlyGluValThr 214
b 479 GTTCAGGTCAATTGACAGCACACCGTGCATGCAATGCAGACGATGCGTACAGGGGGAAGTCACC 420
Y 215 GluLysMetMetCysAlaGlyIleProGluGlyValAspThrCysGlnGlyAspSer 234
b 419 GAGAAGATGATGTGTGCAGGCATCCCGGAAGGGGTGTGGACACCTGCCAGGGTGACAGT 360
Y 235 GlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIleValSerTrpGly 254
b 359 GGTGGGCCCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGCATCGTTAGCTGGGC 300
Y 255 TyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsn 274
b 299 TATGGCTGCGGGGCCCCGAGCACCCCCAGGAGTATACACCAAGGTCTCAGCCTATCTCACC 240
Y 275 TrpIleTyrAsnValTrpLys 281
b 239 TGGATCTACATGTCTGTGAAG 219

RESULT 9
I924182/c
OCUS
EFINITION
wn53f11.x1 NCI CGAP Lul9 Homo sapiens cDNA clone IMAGE:2449197 3',
similar to TR:060235 AIRWAY TRYPSIN-LIKE PROTEASE. ;, mRNA
sequence.
AI924182 648 bp mRNA linear EST 02-SEP-1999
AI924182
AI924182
AI924182.1 GI:5660146
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 648)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/image.html at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 455.
Location/Qualifiers
1. .648
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2449197"
/tissue type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev stage="adult"
/lab_host="DH10B (phage-resistant)"
/clone lib="NCI CGAP Lul9"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo. "

FEATURES
source
BM773306 563 bp mRNA linear EST 04-MAR-2002
K-EST0057738 S3SNU16s1 Homo sapiens cDNA clone S3SNU16s1-17-C08 5',
mRNA sequence.
BM773306
BM773306.1 GI:19102921
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 563)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 17 row: C column: 08
High quality sequence stop: 563.
Location/Qualifiers

BASE COUNT 144 a 174 c 180 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 2.68e-81 Length: 648
Score: 808.00 Matches: 146
Percent Similarity: 99.32% Conservative: 0
Best Local Similarity: 99.32% Mismatches: 1
Query Match: 50.50% Indels: 0
DB: 9 Gaps: 0

US-09-607-745-9 (1-292) x AI924182 (1-648)

QY 135 AspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArg 154
Db 648 GACAATGACATGCCCTCATGAAGCTGCAGTTCCTCCACTCACTTCTCAGGCACAGTCAGG 589
QY 155 ProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrProLeuTrpIleIle 174
Db 588 CCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCAAGCCACCCCACTCTGGATCATT 529
QY 175 GlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSer 194
Db 528 GGATGGGGCTTTACGAAGCAGAATGGAGGAAGATGTCTGACATACTGTCAGGCGTCA 469
QY 195 ValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGlnGlyGluValThr 214
Db 468 GTCCAGGTCAATTGACAGCACACGGTGTCAATGCAGACGATCGTACCAGGGGGAATCACC 409
QY 215 GluLysMetMetCysAlaGlyIleProGluGlyValAspThrCysGlnGlyAspSer 234
Db 408 GAGAAGATGATGTGTGCAGGCATCCCGGAAGGGGTGTGGACACCTGCCAGGGTGACAGT 349
QY 235 GlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIleValSerTrpGly 254
Db 348 GGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGCATCGTTAGCTGGGGC 289
QY 255 TyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsn 274
Db 288 TATGGCTGCGGGGCCCCGAGCACCCCCAGGAGTATACACCAAGGTCTCAGCCTATCTCAAC 229
QY 275 TrpIleTyrAsnValTrpLys 281
Db 228 TGGATCTACATGTCTGTGAAG 208

RESULT 10
BM773306

LOCUS

DEFINITION
K-EST0057738 S3SNU16s1 Homo sapiens cDNA clone S3SNU16s1-17-C08 5',
mRNA sequence.

ACCESSION
BM773306

VERSION
BM773306.1

KEYWORDS
EST.

SOURCE
Homo sapiens (human)

ORGANISM
Homo sapiens

REFERENCE
1 (bases 1 to 563)

AUTHORS
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.

TITLE
21C Frontier Korean EST Project 2001

JOURNAL
Unpublished

COMMENT
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 17 row: C column: 08
High quality sequence stop: 563.
Location/Qualifiers

FEATURES
Location/Qualifiers

```
source
1. .563
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S3SNU16s1-17-C08"
/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblast-like"
/cell_line="SNU-16"
/lab_host="Top10F"
/clone_lib="S3SNU16s1"
/note="Organ: Stomach; Vector: pTZ18RP1; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dT-selected mRNA by
priming with dT-tailed vector. The dT-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 ~ 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and N(dT)14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNAs prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the subtracted cDNA
libraries were constructed by transformaion of the
remaining DNA into competent cells E. coli Top10F' with
electroporation method."
```

BASE COUNT 121 a 172 C 148 G 122 t

ALIGNMENT Scores:

red. No.:	3.84e-80	Length:	563
core:	797.00	Matches:	151
Percent Similarity:	75.50%	Conservative:	0
est Local Similarity:	75.50%	Mismatches:	0
Query Match:	49.81%	Indels:	49
B:	12	Gaps:	1

5-09-607-745-9 (1-292) x BM773306 (1-563)

86 TrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysVal 105

2 TGGGTCTCTACGGCAGCCCACTGTCTCAGGAACATACCGATGTGTTCAACTGGAAGGTG 61

106 ArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIle 125

62 CGGGCAGGCTCAGACAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATC 121

126 IleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhe 145

122 ATTGAATTCAACCCCATGTACCCCAAAGACAATGACATCGCCCTCATGAAGTGCAGTTC 181

146 ProLeuThrPheSerGlyThrValArgProIleCysLeuProphePheAspGluGluLeu 165

182 CCACTCACTTCTCAGGCACAGTCAGGCCCATCTGTGTGCCCTTCTTTGATGAGGAGCTC 241

166 ThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLys 185

242 ACTCCAGCCACCCCACTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAAATGGA----- 295

```
QY 186 MetSerAspIleLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAla 205
Db 295 -----
QY 206 AspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGly 225
Db 295 -----
QY 226 GlyValAspThrCysGlnGly-AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTr 245
Db 296 -----GGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCACTGACCACTG 337
QY 245 pHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyVa 265
Db 338 GCATGTGGTGGGCATCGTTAGTTGGGCTATGGCTGCGGGGGCCGAGCACCCAGAGT 397
QY 265 lTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 398 ATACACCAAGTCTCAGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 455

RESULT 11
BM746729
LOCUS
DEFINITION
K-EST0021071 S3SNU16 Homo sapiens cDNA clone S3SNU16-19-D01 5',
mRNA sequence.
ACCESSION
BM746729
VERSION
BM746729.1 GI:19076313
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 572)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 19 row: D column: 01
High quality sequence stop: 572.
Location/Qualifiers
1. .572
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S3SNU16-19-D01"
/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblast-like"
/cell_line="SNU-16"
/lab_host="Top10F"
/clone_lib="S3SNU16"
/note="Organ: Stomach; Vector: pTZ18RP1; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dT-selected mRNA by
priming with dT-tailed vector. The dT-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F' by electroporation method."
```

The cDNA libraries constructed by this method are full-length enriched cDNA library."

ASE COUNT	123 a	176 c	149 g	124 t
RIGIN				
Alignment Scores:				
red. No.:	3.94e-80	Length:	572	
core:	797.00	Matches:	151	
ercent Similarity:	75.50%	Conservative:	0	
est Local Similarity:	75.50%	Mismatches:	0	
uery Match:	49.81%	Indels:	49	
B:	12	Gaps:	1	
S-09-607-745-9 (1-292) x BM746729 (1-572)				
Y	86	TrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysVal	105	
b	2	TGGGTCTTCACGGCAGCCCACTGCTTCAGGAACATACCGATGTGTCAACTGGAAGGTG	61	
Y	106	ArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIle	125	
b	62	CGGGCAGGCTCAGACAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATC	121	
Y	126	IleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhe	145	
b	122	ATTGAATTCAACCCCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTC	181	
Y	146	ProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeu	165	
b	182	CCACTCACTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTCTTTGATGAGGAGCTC	241	
Y	166	ThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLys	185	
b	242	ACTCCAGCCACCCCACTCTGGATCATTTGGATGGGGCTTTACGAAGCAGATGGA-----	295	
Y	186	MetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAla	205	
b	295	-----	295	
Y	206	AspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGly	225	
b	295	-----	295	
Y	226	GlyValAspThrCysGlnGly-AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTr	245	
b	296	-----GGGTGACAGTGTGGGGCCCTGTGTACCAATCTGACCAAGTG	337	
Y	245	pHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyVa	265	
b	338	GCATGTGGTGGCATCGTTAGTTGGGGCTATGGCTCGGGGGCCGAGCACCCCGAGAGT	397	
Y	265	lTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	284	
b	398	ATACACCAAGGCTCTCAGCCTATCTCACTGGATCTACAATGTCTGGAAGGCTGAGCTG	455	
RESULT 12				
W747250				
OCUS	BM747250	578 bp	mRNA	linear
EFINITION	K-EST0021773 S3SNU16s1	Homo sapiens	cDNA clone	S3SNU16s1-2-A03 5',
	mRNA sequence.			
CCSESSION	BM747250			
ERSION	BM747250.1	GI:19076856		
EYWORDS	EST.			
OURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 578)			
AUTHORS	Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,			
	Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and			
	Kim,Y.S.			
TITLE	21C Frontier Korean EST Project 2001			
JOURNAL	Unpublished			

COMMENT

Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 2 row: A column: 03
High quality sequence stop: 578.

FEATURES

source

1. .578
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S3SNU16s1-2-A03"
/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblast-like"
/cell_line="SNU-16"
/lab_host="Top10F"
/clone_lib="S3SNU16s1"
/note="Organ: Stomach; Vector: pTZ18RP1; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dT-selected mRNA by
priming with dT-tailed vector. The dT-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 ~ 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and N(dT)14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNAs prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the subtracted cDNA
libraries were constructed by transformaion of the
remaining DNA into competent cells E. coli Top10F' with
electroporation method."

BASE COUNT 124 a 179 c 150 g 125 t

ORIGIN

Alignment Scores:

Pred. No.: 4e-80 Length: 578
Score: 797.00 Matches: 151
Percent Similarity: 75.50% Conservative: 0
Best Local Similarity: 75.50% Mismatches: 0
Query Match: 49.81% Indels: 49
DB: 12 Gaps: 1

US-09-607-745-9 (1-292) x BM747250 (1-578)

QY	86	TrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysVal	105
Db	2	TGGGTCTTCACGGCAGCCCACTGCTTCAGGAACATACCGATGTGTCAACTGGAAGGTG	61
QY	106	ArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIle	125
Db	62	CGGGCAGGCTCAGACAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATC	121
QY	126	IleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhe	145

122 ATTGAATTCACCCCATGTACCCCAAGACAAATGACATGCCCTCATGAAGCTGCAGTTC 181
146 ProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeu 165
182 CCACTCACTTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGATGAGGAGCTC 241
166 ThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLys 185
242 ACTCCAGCCACCCCACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGA----- 295
186 MetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAla 205
295 ----- 295
206 AspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGly 225
295 ----- 295
226 GlyValAspThrCysGlnGly-AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTr 245
296 -----GGGTGACAGTGTGGGCCCCCTGTATGATGACCAATCTGACCAAGTG 337
245 pHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyVa 265
338 GCATGTGGTGGGCATCGTTAGTTGGGCTATGGCTGCGGGGGCCGAGCACCCAGGAGT 397
265 lTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
398 ATACACCAAGGTCTCAGCCTATCTCACTGGATCTACAATGCTGGAAGGCTGAGCTG 455

RESULT 13
BM788452
LOCUS
DEFINITION
K-EST0067604 S3SNU16 Homo sapiens cDNA clone S3SNU16-36-A11 5',
mRNA sequence.
BM788452
BM788452.1 GI:19136684
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 593)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 36 row: A column: 11
High quality sequence stop: 593.
Location/Qualifiers
1. 593
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S3SNU16-36-A11"
/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblast-like"
/cell_line="SNU-16"
/lab_host="T0p10P"
/clone_lib="S3SNU16"
/note="Organ: Stomach; Vector: pTZ18RP1; Site 1: EcoRI;
Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped

intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dT-selected mRNA by
priming with dT-tailed vector. The dT-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT 126 a 183 c 154 g 130 t
ORIGIN

Alignment Scores:
Pred. No.: 4.16e-80 Length: 593
Score: 797.00 Matches: 151
Percent Similarity: 75.50% Conservative: 0
Best Local Similarity: 75.50% Mismatches: 0
Query Match: 49.81% Indels: 49
DB: 12 Gaps: 1

US-09-607-745-9 (1-292) x BM788452 (1-593)

QY 86 TrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysVal 105
Db 2 TGGGTCTCTCAGGCAGCCCACTGCTTCAGGAACATATACCATGTGTTCAACTGGAAGGTG 61
QY 106 ArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIle 125
Db 62 CGGGCAGGCTCAGACAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAGATCATCATC 121
QY 126 IleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhe 145
Db 122 ATTGAATTCAACCCCATGTACCCCAAGACAAATGACATGCGCCCTCATGAAGCTGCAGTTC 181
QY 146 ProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeu 165
Db 182 CCACTCACTTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTC 241
QY 166 ThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLys 185
Db 242 ACTCCAGCCACCCCACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGA----- 295
QY 186 MetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAla 205
Db 295 ----- 295
QY 206 AspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGly 225
Db 295 ----- 295
QY 226 GlyValAspThrCysGlnGly-AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTr 245
Db 296 -----GGGTGACAGTGTGGGCCCCCTGTATGATGACCAATCTGACCAAGTG 337
QY 245 pHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyVa 265
Db 338 GCATGTGGTGGGCATCGTTAGTTGGGCTATGGCTGCGGGGGCCGAGCACCCAGGAGT 397
QY 265 lTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 398 ATACACCAAGGTCTCAGCCTATCTCACTGGATCTACAATGCTGGAAGGCTGAGCTG 455

RESULT 14
BM772743
LOCUS
DEFINITION
K-EST0056963 S3SNU16s1 Homo sapiens cDNA clone S3SNU16s1-22-C01 5',
mRNA sequence.
BM772743
BM772743.1 GI:19102358
EST.
KEYWORDS

BM772743 596 bp mRNA linear EST 04-MAR-2002
K-EST0056963 S3SNU16s1 Homo sapiens cDNA clone S3SNU16s1-22-C01 5',
mRNA sequence.

```

SOURCE
ORGANISM      Homo sapiens (human)
              Homo sapiens
REFERENCE     1 (bases 1 to 596)
AUTHORS      Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
              Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
              Kim,Y.S.
TITLE        21C Frontier Korean EST Project 2001
JOURNAL      Unpublished
COMMENT      Contact: Kim YS
              Genome Research Center
              Korea Research Institute of Bioscience & Biotechnology
              52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
              Tel: +82-42-860-4470
              Fax: +82-42-860-4409
              Email: yongsung@mail.kribb.re.kr
              Plate: 22 row: C column: 01
              High quality sequence stop: 596.
              Location/Qualifiers
                source
                  1..596
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="S3SNU16s1-22-C01"
                    /sex="F"
                    /tissue_type="Ascites"
                    /cell_type="Lymphoblast-like"
                    /cell_line="SNU-16"
                    /lab_host="Top10F"
                    /clone_lib="S3SNU16s1"
                    /note="Organ: Stomach; Vector: pTZ18RP1; Site: 1: EcoRI;
                    Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
                    bacterial alkaline phosphatase (BAP) and then decapped
                    with tobacco acid pyrophosphatase (TAP). The decapped
                    intact mRNA was ligated with DNA-RNA linker including EcoR
                    I site by treatment of T4 RNA ligase and the first strand
                    cDNA was synthesized from oligo dT-selected mRNA by
                    priming with dT-tailed vector. The dT-tailed vector was
                    adjusted to have about 60nt. The cDNA vector was
                    circularized with E. coli DNA ligase after digestion of
                    EcoRI which site is also included in vector. An RNA strand
                    converted to a DNA strand by Okayama-Berg method. The
                    obtained cDNA vectors were used for transformation of
                    competent cells E. coli Top10F, by electroporation method.
                    The cDNA libraries constructed by this method are
                    full-length enriched cDNA library. After analyzing and
                    sequencing about 2,000 ~ 3,000 colonies in original cDNA
                    library, the abundant cDNAs were selected and amplified by
                    PCR reaction using vector region primer including T7
                    promoter as 5' primer and N(dT)14 as 3' primer. The PCR
                    products were used as template for synthesis of
                    biotinylated single stranded RNA by in vitro transcription
                    reaction. The synthesized RNA probes were hybridized with
                    antisense single stranded cDNAs prepared from original
                    library and incubated with avidin-gel. After removing
                    DNA-RNA hybrids by centrifuge, the subtracted cDNA
                    libraries were constructed by transfection of the
                    remaining DNA into competent cells E. coli Top10F with
                    electroporation method."
              ASSEMBLY COUNT 127 a 184 C 155 G 130 t
              RIGIN

Alignment Scores:
red. No.:      4.19e-80      Length:      596
core:          797.00      Matches:      151
Percent Similarity: 75.50%      Conservative: 0
Best Local Similarity: 75.50%      Mismatches: 0
Query Match:    49.81%      Indels:      49
3:             12          Gaps:         1

5-09-607-745-9 (1-292) x BM772743 (1-596)

86 TrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysVal 105
|||
2 TGGGTCTCTACGGCAGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTG 61
|||
106 ArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIle 125
|||
62 CGGGCAGGCTCAGACAACTGGGCAGCTTCCCATCCCTGGTGTGGCCAGATCATCATC 121
|||
126 IleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhe 145
|||
122 ATTGAATTCAACCCCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTC 181
|||
146 ProLeuThrPheSerGlyThrValArgProIleCysLeuProphePheAspGluGluLeu 165
|||
182 CCACTCACTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTC 241
|||
166 ThrProAlaThrProLeuTrpIlelleGlyTrpGlyPheThrLysGlnAsnGlyGlyLys 185
|||
242 ACTCCAGCCACCCCACTCTGGATCATTTGATGGGCTTTACGAAGCAGAAATGGA----- 295
|||
186 MetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAla 205
|||
295 ----- 295
|||
206 AspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGly 225
|||
295 ----- 295
|||
226 GlyValAspThrCysGlnGly-AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTr 245
|||
296 -----GGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAGTG 337
|||
245 pHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyVa 265
|||
338 GCATGTGTGGGCATCGTTAGTGGGGCTATGGCTGCGGGGGCCCGAGCACCCAGGAGT 397
|||
265 lTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
|||
398 ATACACCAAGGTCTCAGCCTATCTCAACTGGATCTACAATGTCTGTGAAGGCTGAGCTG 455
|||

RESULT 15
BM788163
LOCUS        BM788163
DEFINITION   K-EST0067267 S3SNU16 Homo sapiens cDNA clone S3SNU16-35-E10 5',
              mRNA sequence.
ACCESSION    BM788163
VERSION      BM788163.1 GI:19136395
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 614)
AUTHORS      Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
              Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
              Kim,Y.S.
TITLE        21C Frontier Korean EST Project 2001
JOURNAL      Unpublished
COMMENT      Contact: Kim YS
              Genome Research Center
              Korea Research Institute of Bioscience & Biotechnology
              52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
              Tel: +82-42-860-4470
              Fax: +82-42-860-4409
              Email: yongsung@mail.kribb.re.kr
              Plate: 35 row: E column: 10
              High quality sequence stop: 614.
              Location/Qualifiers
                source
                  1..614
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="S3SNU16-35-E10"

```

Db 398 ATACACCAAGGTCTCAGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 455

Search completed: November 23, 2003, 12:06:05
Job time : 2086.73 secs

/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblast-like"
/cell_line="SNU-16"
/lab_host="Top10F"
/clone_lib="S3SNU16"
/note="Organ: Stomach; Vector: pTZ18RP1; Site 1: EcoRI;
Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

SE COUNT 132 a 189 c 159 g 134 t
IGIN

ignment Scores:

ed. No.: 4.39e-80 Length: 614
ore: 797.00 Matches: 151
rcent Similarity: 75.50% Conservative: 0
st Local Similarity: 75.50% Mismatches: 0
ery Match: 49.81% Indels: 49
12 Gaps: 1

-09-607-745-9 (1-292) x BM788163 (1-614)

86 TrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysVal 105
2 TGGGTCTCTACGGCAGCCCACTGCTTCAGGAACAATACCGATGTGTTCAACTGGAAGGTG 61
106 ArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIle 125
62 CGGGCAGGCTCAGACAAACTGGGCAGCTTCCCATCGCTGTGGCCAAGATCATCATC 121
126 IleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhe 145
122 APTGAATTCAACCCCATGTATACCCCAAGACAATGACATGCGCCTCATGAAGCTGCAGTTC 181
146 ProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeu 165
182 CCACTCACTTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTC 241
166 ThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLys 185
242 ACTCCAGCCACCCCACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAATGGA----- 295
186 MetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAla 205
295 ----- 295
206 AspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGly 225
295 ----- 295
226 GlyValAspThrCysGlnGly-AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTr 245
296 -----GGGTGACAGTGGTGGCCCCCTGATGTACCAATCTGACCAGTG 337
245 pHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyVa 265
338 GCATGTGGTGGGCATCGTTAGTTGGGGCTATGGCTGCGGGGGCCCCGACACCCAGGAGT 397
265 lTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

Run on: November 23, 2003, 07:49:55 ; Search time 391.919 Seconds
(without alignments)
2996.170 Million cell updates/sec

File: US-09-607-745-2

Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVPLRKP.....VYTKVSAYLWYINWKAEL 435

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

MODEL=frame+p2n.model -DEV=xlh
Q=/cgn2_1/USPTO spoel/US09607745/runat 21112003 144344 22176/app query.fasta_1.1038
DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USRR=US09607745 @CGN 1 1 401 @runat 21112003 144344 22176 -NCPU=6 -ICPU=3
NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
- 25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2342	100.0	1305	22	AAD02556 Human seripancrin
2	2338	99.8	2038	20	AAX87154 Human protease HUP
3	2337	99.8	1314	25	ABX76354 Lung cancer-associ
4	2337	99.8	2081	24	AAI72976 CJA8 preferred cDN
5	2337	99.8	2307	24	ABX76529 cDNA encoding huma
6	2335	99.7	2121	24	AAI64284 Human serine prote
7	2329	99.4	1479	22	AAD02557 Human seripancrin
8	2329	99.4	2165	24	ABZ35523 Human gene express
9	2324	99.2	2070	21	AAZ90471 Cancer specific ge
10	2324	99.2	2079	22	AAD13169 Human transmembran
11	2324	99.2	2079	24	AAI72975 CJA8 cDNA. Homo s
12	2324	99.2	2079	25	ABSS7763 cDNA encoding huma
13	2319	99.0	2137	22	AAD13114 Human membrane-typ
14	2297.5	98.1	2063	21	AAA37099 Human PRO1570 (UNQ
15	2297.5	98.1	2063	22	AAS46089 Human DNA encoding
16	2297.5	98.1	2063	22	AAF92113 Human PRO1570 cDNA
17	2297.5	98.1	2063	22	AAF54396 DNA encoding prote
18	2297.5	98.1	2063	24	ABX74433 Human cDNA encodin
19	2297.5	98.1	2063	24	ABK11090 cDNA encoding tumo
20	2297.5	98.1	2063	25	ACA57847 Human PRO1570 cDNA
21	2297.5	98.1	2063	25	ACA58865 cDNA encoding huma
22	2297.5	98.1	2063	25	ACA60418 Novel human secret
23	2297.5	98.1	2063	25	ACA63428 cDNA encoding huma
24	2297.5	98.1	2063	25	ABX98317 Human cDNA encodin
25	2297.5	98.1	2063	25	ABX98819 Novel human secret
26	2297.5	98.1	2063	25	ACA05864 Human secreted/tra
27	2297.5	98.1	2063	25	ABX97908 Human PRO polynucl
28	2297.5	98.1	2063	25	ABX78692 Human CDNA encodin
29	2297.5	98.1	2063	25	ABX75705 Human PRO polynucl
30	2297.5	98.1	2063	25	ABX76910 Human CDNA encodin
31	2297.5	98.1	2063	25	ABX16750 Human CDNA encodin
32	2251	96.1	1795	22	AAS26880 Human CDNA encodin
33	1668	71.2	1854	22	AAH99574 Human protein enco
34	1287	55.0	699	18	AAT79127 Human serine prote
35	1286	54.9	1281	21	AAA97361 Human colorectal c
36	1259	53.8	1189	24	AAI64290 Protease D-G catay
37	1091	46.6	649	22	AAH33477 Human colon cancer
38	1090	46.5	867	22	AAS30210 DNA encoding rena
39	1090	46.5	867	22	AAS26952 Human CDNA encodin
40	1090	46.5	867	22	AAS33181 DNA encoding human
41	918	39.2	1084	22	AAH34953 Human colon cancer
42	806	34.4	1008	20	AAX04381 Human secreted pro
43	699.5	29.9	2067	22	AAD05797 Human transmembran
44	698.5	29.8	1434	25	ABZ22864 Human dendritic ce
45	697.5	29.8	1314	22	AAF83971 Human SER6 nucleot

ALIGNMENTS

RESULT 1

AAD02556

ID AAD02556 standard; cDNA; 1305 BP.

XX AAD02556;

AC AAD02556;

XX 02-MAY-2001 (first entry)

DT Human seripancrin cDNA.

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

X H Key Location/Qualifiers
T T CDS 1..1305
T T /*tag= a
T T /product= "Human seripancrin protein"
T T /note= "The coding region does not include stop codon"
T T /partial
N WO200104141-A2.
X
X
X D 18-JAN-2001.
X
X F 04-JUL-2000; 2000WO-EP06211.
X
X R 12-JUL-1999; 99EP-0113428.
X
X (MERE) MERCK PATENT GMBH.
X A
X Suendermann B, Hofmann U, Matzku S, Wilbert O;
X
X WPI; 2001-147177/15.
R P-PSDB; AAY72558.
R
X
T New extracellular serine protease Seripancrin, useful for treating
T cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease,
T diabetes, inflammatory disorders, stroke, angiogenesis and aberrant
T wound healing -
X
S Claim 5; Page 37-39; 45pp; English.
X
C The present invention relates to seripancrin polynucleotides, and
C polypeptides encoded by them. Seripancrin are members of serine
C protease family. This protein contains a transmembrane domain,
C a low density lipoprotein (LDL) domain, protease domain and a
C scavenger receptor cysteine-rich (SRCR) domain. The LDL and SRCR
C domains help to define the specificity of seripancrin's intra and
C intermolecular interactions. The polynucleotides and polypeptides
C of the invention are useful for treating and diagnosing diseases
C such as arthritis, chronic obstructive pulmonary disorder (COPD),
C cancer, osteoporosis, aberrant wound healing, angiogenesis,
C inflammatory disorders, diabetes, stroke and cardiovascular diseases.
C Seripancrin genes are useful in chromosome localisation studies,
C as tools for tissue expression studies and also in gene therapy.
C The polypeptides of the invention are used for identifying agonists
C and antagonists useful for treating conditions associated with
C seripancrin imbalance. These polypeptides are also useful as vaccines.
C The present sequence is a cDNA coding for seripancrin protein. The
C seripancrin gene is located on human chromosome 11q22-q23.
X
Q Sequence 1305 BP; 293 A; 368 C; 372 G; 272 T; 0 other;

Alignment Scores:
red. No.: 6.33e-210 Length: 1305
core: 2342.00 Matches: 435
percent Similarity: 100.00% Conservative: 0
est Local Similarity: 100.00% Mismatches: 0
query Match: 100.00% Indels: 0
B: 22 Gaps: 0

S-09-607-745-2 (1-435) x AAD02556 (1-1305)

Y 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
b 1 ATGGATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCTCAAAACCCCTGCGCAACCC 60
Y 21 ArgIleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSer 40
b 61 CGTATCCCCCATGGAGACCTTCAGAAAGGTGGGATCCCCATCATCATAGCACTACTGAGC 120
Y 41 LeuAlaSerIleIleIleValValValIleLysValIleLeuAspLysTyrPhe 60
b 121 CTGGCGAGTATCATATTGTGGTTTCTCATCAAGGTGATTTCTGGATAATACTACTTC 180

QY 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
Db 181 CTCTGCGGGCAGCTCTCCACTTCATCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGAC 240
QY 81 CysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaVal 100
Db 241 TGTCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCCCTGCAGTG 300
QY 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
Db 301 GCAGTCCGCTCTCCAAGGACCGATCCACACTGCAGGTGCTGGACTCGGCCACAGGGGAC 360
QY 121 TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
Db 361 TGGTTCTCTGCTGCTTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCCTGTAGGCAG 420
QY 141 MetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
Db 421 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCCCAGACCAGGATCTG 480
QY 161 AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180
Db 481 GATGTTGTTGAATCACAGAAAACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCC 540
QY 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
Db 541 TGTCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCCTGTGGGAGAGCCTGAAGACC 600
QY 201 ProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
Db 601 CCCCCTGTGGTGGTGGGAGGAGGCCCTCTGTGGATTCTTGGCCTTGGCAGGTGAGCATC 660
QY 221 GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
Db 661 CAGTACGACAAACAGCACAGCTCTGTGAGGGAGCATCTCTGACCCCCCCTGCTGGTCTCTCAGG 720
QY 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260
Db 721 GCAGCCCCACTGCTTCAGGAAACATACCGATGTGTTCACTGGAAGGTGCGGGCAGGCTCA 780
QY 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsn 280
Db 781 GACAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATATTGAATCAAC 840
QY 281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
Db 841 CCCATGTACCCCAAGACATGACATGCCCTCATGAAGTGCAGTTCCTCCACTCACTTTC 900
QY 301 SerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThr 320
Db 901 TCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGTCACTCCAGCCACC 960
QY 321 ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle 340
Db 961 CCACCTCTGGATCATTTGGATGGGCTTTTACGAAGCAGATGGAGGAAGATGTCTGACATA 1020
QY 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyr 360
Db 1021 CTGCTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGTCAATGCAGACGATCGGTAC 1080
QY 361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr 380
Db 1081 CAGGGGGAAGTCAACCGAGAAGATGATGTGTGACGATCCCGGAAGGGGTGTGGACACC 1140
QY 381 CysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGly 400
Db 1141 TGCCAGGTTGACAGTGTGGGCCCTCTGATGTACCAATCTGACCAGTGGCATGTGTGGTGGGC 1200
QY 401 IleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValThrLysVal 420
Db 1201 ATCGTAGCTGGGGCTATGGCTGCGGGGCCCGAGCACCCAGGAGTATACACCAAGGTC 1260
QY 421 SerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435

1261 TCAGCCTATCTCAACTGGATCTACAATGTCGGAAGGCTGAGCTG 1305

RESULT 2

AX87154

D AAX87154 standard; cDNA; 2038 BP.

X X

C AAX87154;

T 27-SEP-1999 (first entry)

X Human protease HUPM-6 cDNA.

E Serine protease; human; HUPM-6; cell proliferation; cancer;

W immune disorder; inflammation; therapy; ss.

W Homo sapiens.

X Key

X Location/Qualifiers

T CDS

T 200..1507

X /*tag= a

X WO9936550-A2.

X 22-JUL-1999.

X 12-JAN-1999; 99WO-US000655.

X 16-JAN-1998; 98US-0008271.

X (INCY-) INCYTE PHARM INC.

X Bandman O, Corley NC, Guegler KJ, Hillman JL, Shah P;

X Tang YT, Yue H;

X WPI; 1999-430616/36.

X P-PSDB; AAY06437.

X Novel human protease molecules useful in the treatment of

X developmental disorders and/or cancers

X Claim 8; Page 86-87; 90pp; English.

X This nucleotide sequence codes for HUPM-6 (see AAY06437), a novel

X human protease. HUPM-6 cDNA was initially identified in Incyte

X Clone 1337018 from the colon cDNA library COLN0713 using a

X computer search for amino acid sequence alignments. The present

X sequence is a consensus sequence derived from overlapping and/or

X extended nucleic acid sequences: Incyte Clones 1271725 (TESTTUT02),

X 1337018, 586592 and 588598 (UTRSNOT01). A fragment comprising

X nucleotides 900-949 of the present sequence can be used for

X hybridisation. This sequence encompasses an active site residue.

X Northern analysis shows expression of HUPM-6 in gastrointestinal,

X and male an female reproductive cDNA libraries. Approximately 65%

X of these libraries are associated with neoplastic disorders, and

X 22% with the immune response. The invention provides 12 new human

X proteases, i.e. HUPM-1 to -12 (see AAY06432-43), and the

X polynucleotides encoding them (see AAX87149-60). Also provided are

X vectors, host cells and methods for producing HUPM polypeptides, as

X well as agonists and antagonists of HUPM. Methods for treating or

X preventing cell proliferative disorders and immune disorders using

X HUPM or HUPM antagonists are claimed.

X Sequence 2038 BP; 462 A; 591 C; 569 G; 416 T; 0 other;

X Alignment Scores:

X ed. No.:

X ore:

X rcent Similarity:

X st Local Similarity:

X ery Match:

X :

X 2.72e-209

X 2338.00

X 100.00%

X 99.77%

X 99.83%

X 20

X Length:

X Matches:

X Conservative:

X Mismatches:

X Indels:

X Gaps:

X 2038

X 434

X 1

X 0

X 0

X 0

X 0

X 0

X 0

X 0

US-09-607-745-2 (1-435) x AAX87154 (1-2038)

QY 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
 Db 200 ATGGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAACCC 259
 QY 21 ArgIleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSer 40
 Db 260 CGTATCCCCATGGAGACCTTCAGAAAGGTGGGATCCCCATCATCATAGCACTACTGAGC 319
 QY 41 LeuAlaSerIleIleValValValLeuIleLysValIleLeuAspLysTyrTrpPhe 60
 Db 320 CTGGCGAGTATCATCATTTGTTGTTCTCATCAAGGTGATTCTGGATAAATACTACTTC 379
 QY 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
 Db 380 CTCTGCGGGCAGCTCTCCACTTCATCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGAC 439
 QY 81 CysProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaVal 100
 Db 440 TGTCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCCTGCAGTG 499
 QY 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
 Db 500 GCAGTCCGCTCTCCAAGGACCGATCCACACTGCAGGTGCTGGACTCGGCCACAGGGAAC 559
 QY 121 TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
 Db 560 TGGTCTCTGCTGCTGTTGACAACTTCACAGAAGCTCTCGCTGAGACAGCTGTAGGAG 619
 QY 141 MetGlyTyrSerSerLysProThrPheArgAlaValGluIleClyProAspGlnAspLeu 160
 Db 620 ATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACAGCATCTG 679
 QY 161 AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180
 Db 680 GATGTTGTTGAAATCACAGAAACACAGCAGGAGCTTTCGCATGCGGAACTCAAGTGGG 739
 QY 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
 Db 740 TGTCTCTCAGGCTCCCTGGTCTCCCTGCAGCTGTCTTGCCTGTGGGAGAGCCCTGAAGAC 799
 QY 201 ProArgValValGlyGlyGluGluAlaSerValAspSerTrpTrpGlnValSerIle 220
 Db 800 CCCCTGTGTGGTGGGAGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGGTTCAGCATC 859
 QY 221 GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
 Db 860 CAGTACGACAAACAGCAGCTGTGTGGAGGAGCATCTTGGACCCCACTGGGTCTCTCAG 919
 QY 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260
 Db 920 GCAGCCCACTGCTTCAGGAACATACCGATGTGTTCACTGGAAGGTGCGGCGAGGCTCA 979
 QY 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsn 280
 Db 980 GACAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATCATGAATTCAAC 1039
 QY 281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
 Db 1040 CCCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGAGTTCCTCCACTCACTTC 1099
 QY 301 SerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThr 320
 Db 1100 TCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGAGCTCACTCCAGCCACC 1159
 QY 321 ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle 340
 Db 1160 CCACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATA 1219
 QY 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyr 360
 Db 1220 CTGCTGCAGGCGTCACTCCAGGTCAATTGACAGCACACCGGTGCAATGCAGACGATGCGTAC 1279

Y 361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThr 380
b 1280 CAGGGGAAGTACCCGAGAAGATGATGTGTGACGCATCCCGAAGGGGTGTGGACACC 1339
Y 381 CysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValGly 400
b 1340 TGCCAGGGTGACAGTGTGGGCCCTGTGATGTACCAATCTGACCAGTGGCATGTGGTGGC 1399
Y 401 IleValSerTrpGlyTyrGlyCysGlyProSerThrProGlyValTyrThrLysVal 420
b 1400 ATCGTTAGCTGGGGCTATGGCTGCGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTC 1459
Y 421 SerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
b 1460 TCAGCCTATCTCAACTGGATCTACAATGTCCTGGAAGGCTGAGCTG 1504

RESULT 3
BX76354
D ABX76354 standard; DNA; 1314 BP.
C ABX76354;
X 02-APR-2003 (first entry)
X Lung cancer-associated polynucleotide #218.
E Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;
W antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
W small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
W chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
W interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.

S Unidentified.
X WO200286443-A2.
X 31-OCT-2002.
X 18-APR-2002; 2002WO-US12476.
X 18-APR-2001; 2001US-284770P.
R 10-MAY-2001; 2001US-290492P.
R 09-NOV-2001; 2001US-339245P.
R 13-NOV-2001; 2001US-350666P.
R 29-NOV-2001; 2001US-334370P.
R 12-APR-2002; 2002US-372246P.

(EOSB-) EOS BIOTECHNOLOGY INC.
X Aziz N, Murray R;
X WPI; 2003-093161/08.
X P-PSDB; ABUS6625.
T Detecting a lung cancer-associated transcript in a cell from a patient
T for treating lung cancer, by contacting a biological sample from the
T patient with a polynucleotide that exhibits increased or decreased
T expression in lung cancer -
S Claim 22; Page 353; 453pp; English.

X The invention relates to a method for detecting a lung cancer-associated
C transcript in a cell from a patient, comprising contacting a biological
C sample from the patient with a polynucleotide that selectively hybridises
C to a sequence that is at least 80 % identical to a gene that exhibits
C increased or decreased expression in lung cancer samples. Lung
C cancer-associated polynucleotides and polypeptides are used for
C identifying a compound that modulates a lung cancer-associated
C polypeptide, for inhibiting proliferation of a lung cancer-associated
C cell to treat lung cancer in a patient and for treating a mammal having
C lung cancer by administering a modulatory compound identified. The
C methods are useful for treating lung cancer, such as small cell lung

CC cancer, non-small cell lung cancer or other benign or precancerous
CC lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive
CC pulmonary disease, fibrosis, hypersensitivity pneumonitis, interstitial
CC pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides
CC and polypeptides are useful for diagnostic purposes and as targets for
CC screening for therapeutic compounds that modulate lung cancer, such as
CC antibodies. Sequences ABX76124-ABX76474 represent lung cancer-associated
CC polynucleotides of the invention.
XX

SQ Sequence 1314 BP; 297 A; 369 C; 373 G; 275 T; 0 other;

Alignment Scores:
Pred. No.: 1.88e-209 Length: 1314
Score: 2337.00 Matches: 434
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 25 Gaps: 0

US-09-607-745-2 (1-435) x ABX76354 (1-1314)

QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
Db 10 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAACCCCGT 69
QY 22 IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeu 41
Db 70 ATCCCCATGGAGACCTTCAGAAAGTGGGGATCCCCATCATCATAGCACTACTAGCCTG 129
QY 42 AlaSerIleIleIleValValValIleLysValIleLeuAspLysTyrTyrPheLeu 61
Db 130 GCGAGTATCATCATTTGTGTTCTCTCATCAAGGTGATTCTGGATAAATACTACTTCCTC 189
QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
Db 190 TGCGGGCAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTACGGAGAGCTGGACTGT 249
QY 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
Db 250 CCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAGGGCCTGCAGTGGCA 309
QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
Db 310 GTCGGCCTCTCCAAGGACCGATCCACACTGCAGGTGTGACTCGGCCACAGGGAACCTGG 369
QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
Db 370 TTCTCTGCCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGAGATG 429
QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
Db 430 GGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCAGACAGGATCTGGAT 489
QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
Db 490 GTTGTGTAATCACAGAAACACCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCCCTGT 549
QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
Db 550 CTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGGCTGTGGGAAGAGCCTGAAGACCCCC 609
QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
Db 610 CGTGTGTGGTGGGAGGAGGAGCCTCTGTGGATTCTTGGCCTTGGCAGGTCCAGATCCAG 669
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
Db 670 TACGACAAACAGCAGCTCTGTGGAGGGAGCATCTCTGACCCCACTGGTCTCTCAGCGCA 729
QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
Db 730 GCCCACTGCTTCAGGAACATACCGATGTGTTCACACTGGAAGGTGCGGGCAGGCTCAGAC 789

262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
 790 AAACCTGGGAGCTTCCCATCCCTGGCTGGCAAGATCATCATTAATCAACCC 849
 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
 850 ATGTACCCCAAGACATGACATGCCCTCATGAAGCTGCAGTCCCACTCACTTCTCA 909
 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
 910 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGATGAGGAGCTCACTCCAGCCACCCA 969
 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
 970 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGAGATGTCTGACATACTG 1029
 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 361
 1030 CTGCAGGCGTCAGTCCAGGTCACTTGACAGCACACGGTGCATGCGACGATGCGTACCAG 1089
 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThrCys 381
 1090 GCGGAAGTCACCGAGAAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGACACTGC 1149
 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
 1150 CAGGTCAGAGTGTTGGGCCCCCTGATGATCAATCTGACCAAGTGGCATGTGTGGGCATC 1209
 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
 1210 GTTAGCTGGGCTATGGCTCGGGGCGCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1269
 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
 1270 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1311

SULT 4
 I72976

AAI72976 standard; cDNA; 2081 BP.

AAI72976;

21-AUG-2002 (first entry)

CJA8 preferred cDNA.

Gene; colorectal cancer; CGA7; CJA8; modulating protein; screening; drug candidate; vaccine; ss.

Homo sapiens.

Key Location/Qualifiers
 CDS 215..1528
 /*tag= a
 /product= "CJA8"

US2002042067-A1.

11-APR-2002.

08-MAY-2001; 2001US-0851588.

17-AUG-2000; 2000US-0642252.

06-SEP-2000; 2000US-0656002.

(MACK/) MACK D.

(GISH/) GISH K C.

(WILS/) WILSON K E.

Mack D, Gish KC, Wilson KE;

WPI; 2002-453647/48.

P-PSDB; AAG79359.

XX Screening drug candidates for treating colorectal cancer, comprises
 PT determining the effect of the candidate on the expression profile gene
 PT of CGA7 or CJA8
 XX Claim 69; Fig 7; 40pp; English.
 PS
 XX The sequences given in AAI72973-76 encode the colorectal cancer
 CC proteins, CGA7 and CJA8. The CGA7 and CJA8 proteins are colorectal
 CC cancer modulating proteins and have been mapped to chromosomes 2 (CGA7)
 CC and 11 (CJA8). These sequences may be used in the method of the
 CC invention for screening drug candidates. The method comprises adding
 CC a drug candidate to a cell that expresses an expression profile gene
 CC encoding CGA7, CJA8 or fragments and determining the effect of the
 CC drug candidate on the expression of the expression profile gene. The new
 CC methods are used to screen bioactive agents for the ability to bind to
 CC or modulate the activity of CGA7 or CJA8 and evaluate the effect of a
 CC candidate colorectal cancer drug. An antibody to CGA7 or CJA8 can inhibit
 CC the activity of CGA7 or CJA8, respectively, and is used to screen for
 CC an agent that can interfere with the binding of CGA7 or CJA8 to the
 CC antibody. The antibody can be used to treat colorectal cancer. The
 CC antibody or a fragment of it is used to localize a therapeutic group to
 CC a colorectal cancer tissue, where the therapeutic group is a cytotoxic
 CC agent or a radioisotope. Antisense molecules are used to inhibit
 CC colorectal cancer in a cell. Nucleic acid segments encoding CGA7 or
 CC CJA8 are used in a biochip. CGA7, CJA8 or a nucleic acid encoding it
 CC are used to elicit an immune response. CGA7 or CJA8 is used to
 CC determine the prognosis of an individual with colorectal cancer.
 CC Nucleic acid encoding CGA7 or CJA8 can be used in vaccines.
 XX
 SQ Sequence 2081 BP; 484 A; 597 C; 576 G; 424 T; 0 other;

Alignment Scores:

Pred. No.: 3,47e-209 Length: 2081
 Score: 2337.00 Matches: 434
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.79% Indels: 0
 DB: 24 Gaps: 0

US-09-607-745-2 (1-435) x AAI72976 (1-2081)

QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
 DB 224 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAACCCCTGCGAAACCCGT 283
 QY 22 IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeu 41
 DB 284 ATCCCATGGAGACCTTCAGAAAGGTGGGATCCCCATCATCATAGCAGCTACTGAGCCTG 343
 QY 42 AlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
 DB 344 GCGAGTATCATCATGTGTGTCTCTCATCAAGTGATTTCTGGATAAATACTACTTCTC 403
 QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuLysAspGlyGluLeuAspCys 81
 DB 404 TCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 463
 QY 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
 DB 464 CCCTTGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 523
 QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
 DB 524 GTCCGCTCTCCAGGACCGATCCACACTGAGGTGTGGACTCGGCCACAGGGAACCTGG 583
 QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
 DB 584 TTCTCTGCTGTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCCTGTAGGAGATG 643
 QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
 DB 644 GGCTACAGCAGCAACCCACTTTTCAGAGCTGTGGAGATTGGCCACAGGAGTCTGGAT 703

Y 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
b 704 GTTGTGAAATCACAGAAACAGCCAGGAGCTTCGATGCGGAACCTCAAGTGGGCCCTGT 763
Y 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
b 764 CTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGGTGTGGGAAGAGCTGAAGACCC 823
Y 202 ArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
b 824 CGTGTGGTGGGTGGGAGAGGCCCTCTGTGGATTCTTGGCCTTGGCAGGTCAAGTCCAG 883
Y 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
b 884 TAGCACAAACAGCACGCTCTGTGGAGGGAGCATCTCGACCCCCACTGGTCTCACGGCA 943
Y 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
b 944 GCCCACTGCTTCAGGAACATACCCGATGTGTTCAACTGGAAGGTGCGGCGAGGCTCAGAC 1003
Y 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
b 1004 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGTGGCAAGATCATCATCAATTCACACCC 1063
Y 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
b 1064 ATGTACCCCAAGACAAATGACATGACATGCCCTCATGAAGCTGCAGTTCCTCACTTCTCA 1123
Y 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
b 1124 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGTCACTCCAGCCACCCCA 1183
Y 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
b 1184 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATCTG 1243
Y 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 361
b 1244 CTGCAGCGTCAGTCCAGGTCATTGACAGCACACGGTGCATGCAGACGATGCGTACCAG 1303
Y 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
b 1304 GGGGAAGTCACCGAGAAGATGATGTGTGACGATCCCGAAGGGGTGTGGACACCTGC 1363
Y 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
b 1364 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCACTGGATGTGGTGGGCATC 1423
Y 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
b 1424 GTTAGCTGGGCTATGCTGCGGGGCCCGGACACCCAGGAGTATACCAAGGTCTCA 1483
Y 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
b 1484 GCCTATCTCAACTGGATCTACATGTCTGGAAGGCTGAGCTG 1525

RESULT 5

BS76529

D ABS76529 standard; cDNA; 2307 BP.

X X

C ABS76529;

X X

T 11-DEC-2002 (first entry)

X X

E cDNA encoding human ovarian cancer marker OV86.

X X

W Human; ovarian cancer; marker; cancer; familial history; brain disorder;
W central nervous system disorder; bacterial meningitis; viral meningitis;
W Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;
W brain herniation; inflammation; encephalitis; testicular disorder;
W nontuberculous granulomatous orchitis; connective tissue disorder;
W heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;

KW

XX

OS

XX

PN

XX

PD

XX

PF

XX

PR

PR

PR

PR

PR

PR

PR

XX

XX

PI

PI

PI

XX

DR

DR

XX

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

histological type; carcinogenic; ovarian cancer marker; gene; ss.

Homo sapiens.

WO200271928-A2.

19-SEP-2002.

14-MAR-2002; 2002WO-US07826.

14-MAR-2001; 2001US-276025P.

14-MAR-2001; 2001US-276026P.

10-AUG-2001; 2001US-311732P.

19-SEP-2001; 2001US-323580P.

26-SEP-2001; 2001US-324967P.

26-SEP-2001; 2001US-325102P.

26-SEP-2001; 2001US-325149P.

(MILL-) MILLENNIUM PHARM INC.

Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;
Meyers RE, Morrissey MP, Olandt PJ, Sen A, Vieby PO, Mills GB;
Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;

WPI; 2002-723277/78.

P-PSDB; ABG96430.

Assessing whether a patient is afflicted with ovarian cancer, useful in
assessing the stage or progression of the disease, comprises comparing
the expression level of a cancer marker in a sample from a patient and
from a non cancer patient -

Disclosure; Page 438; 481pp; English.

The present invention relates to a new method for assessing whether a
patient is afflicted with ovarian cancer. The method involves comparing
the expression level of a marker in a patient sample and the normal level
of expression of the marker in a control non-ovarian cancer sample, where
the marker is selected from 363 cancer markers described in the
specification. The method of the invention is useful in diagnosing or
characterising cancer, in detecting the presence of cancer as early as
possible, and the recurrence of ovarian cancer. The method may also be of
particular use with patients having an enhanced risk of developing
ovarian cancer (e.g. patients having a familial history of ovarian
cancer). The cancer markers may be used in the management and treatment
of e.g. brain and central nervous system disorders (e.g. bacterial and
viral meningitis, Alzheimer's disease or Parkinson's disease), brain
disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),
inflammations (e.g. bacterial or viral meningitis or encephalitis),
testicular disorders (e.g. nontuberculous granulomatous orchitis),
connective tissue disorders, or heart disorders (e.g. ischaemic heart
disease or atherosclerosis). The compositions and methods may also be
used in assessing the histological type of neoplasm associated with
ovarian cancer, monitoring the progression of ovarian cancer,
determining whether ovarian cancer has metastasized or is likely to
metastasize, selecting a composition for inhibiting ovarian cancer,
assessing the ovarian carcinogenic potential of a compound, or
inhibiting ovarian cancer or at risk of developing ovarian cancer. The
present nucleic acid sequence encodes one of the ovarian cancer markers
described in the invention.

Sequence 2307 BP; 555 A; 647 C; 614 G; 460 T; 31 other;

Alignment Scores:

pred. No.:

3.98e-209

Length:

2307

Score:

2337.00

Matches:

434

Percent Similarity:

100.00%

Conservative:

0

Best Local Similarity:

100.00%

Mismatches:

0

Query Match:

99.79%

Indels:

0

DB:

24

Gaps:

0

US-09-607-745-2 (1-435) x ABS76529 (1-2307)


```

2  AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
284  GATCCTGACAGTGAACACCTCTGAACAGCTCGATGTCAAACCCCTGGCAAAACCCCGT 343
22  IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeu 41
344  ATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATATAGCACTACTAGACCTG 403
42  AlaSerIleIleIleValValLeuIleIleValIleLeuAspLysTyrTyrPheLeu 61
404  GCGAGTATCATCATTTGGTTGTCTCTCATCAAGGTGATTCTGGATAAATACTACTCTC 463
62  CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
464  TCGGGGAGCCCTCTCCACTTTCATCCCGAGAGCAGCTGTGTGACGGAGAGCTGACTGT 523
82  ProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaValAla 101
524  CCTTGGGGGAGGAGGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCCTGCAGTGGCA 583
102  ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
584  GTCCGCTCTCTCAAGGACCATCCACACTGAGGCTGTGGACTGGCCACAGGGAACTGG 643
122  PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
644  TTCTCTGCCTGTTTCGACAACTTCACAGAAAGTCTCTCGTGAGACAGCCTGTAGGCAGATG 703
142  GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
704  GGCTACAGCAGCAAAACCCCTTTCAGAGCTGTGGAGATTGGCCAGACCAAGGATCTGGAT 763
162  ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
764  GTTGTGAATACACAGAAACAGCCAGGAGCTTCGCATCGGAACTCAAGTGGGCCCTGT 823
182  LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
824  CTCTCAGGCTCCCTGGTCTCCCTGCCTGCACTGTCTGTGCTGTGGAGAGAGCCTGAAGACCCC 883
202  ArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
884  CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTCAGCATCCAG 943
222  TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
944  TACGACAAACAGCACGCTCTGTGGAGGAGCATCTCTGACCCCTCCCTGCTCCTCACGGCA 1003
242  AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
1004  GCCCACTGCTTCAGAAACATACCGATGTGTCAACTGGAAGGTGGGAGGCTCAGCTCAGAC 1063
262  LysLeuGlySerPheProSerLeuAlaValAlaIleIleIleIleGluPheAsnPro 281
1064  AAACCTGGGAGCTTCCCATCCCTGGCTGTGGCAAGATCATCATCATTAATTCAAACCC 1123
282  MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
1124  ATGTACCCCAAGACATGACATCGCCCTCATGAAGCTGCAGTCCCACTCACTTTCTCA 1183
302  GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
1184  GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1243
322  LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu 341
1244  CTCTGGATCATTTGGATGGGCTTACGAAGCAGATGGAGGAAGATGTCTGACATATCTG 1303
342  LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 361
1304  CTGCAGGCTCAGTCCAGGTCAATGACAGCACACGGTGCATATGCAGAGATCGGTACCAG 1363
362  GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381

```

```

1364  GCGGAAGTCACCGAAGATGATGTGTGCAGGATCCCGAAGGGGTGTGGACACCTGC 1423
382  GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
1424  CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGTGGGCATC 1483
402  ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
1484  GTTAGCTGGGCTATGGCTGCGGGGCCCGGAGCACCCAGGAGTATACACCAAGGTCTCA 1543
422  AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
1544  GCCTATCTCAACTGGATCTACAAATGTCTGTGAAGGCTGAGCTG 1585

RESULT 6
AA164284
ID  AA164284 standard; cDNA; 2121 BP.
XX
AC  AA164284;
XX
DT  07-MAY-2002 (first entry)
XX
DE  Human serine protease D-G cDNA.
XX
KW  Serine protease; D-G; human; zymogen; enzyme; cytostatic;
KW  antiinflammatory; dermatological; anticoagulation; cancer;
KW  skin disorder; neuropathic pain; inflammatory disorder;
KW  coagulation diathesis; thrombosis; laundry detergent; skin care;
KW  gene therapy; ss.
XX
OS  Homo sapiens.
XX
FH  Key
CDS 277..1584
FT /*tag= a
FT /product= "serine protease D-G"
FT /transl_except= (pos: 1357..1359, aa: Gln)
FT polyA_signal 2100..2105
FT /*tag= b
FT /note= "does not conform to consensus sequence"
XX
XX
XX 10-JAN-2002.
XX
XX 08-JUN-2001; 2001WO-US18568.
XX
XX 30-JUN-2000; 2000US-0607745.
XX
XX (ORTH ) ORTHO-MCNEIL PHARM INC.
XX
XX Darrow AL, Qi J, Andrade-Gordon P;
XX
XX WPI; 2002-106601/14.
XX P-PSDB; AAG78577.
XX
XX Nucleic acid encoding a serine protease called D-G protein which is
XX useful for identifying modulators that are useful for treating a
XX condition which is mediated by protease D-G, e.g. cancer, skin
XX disorders, or neuropathic pain -
XX
XX Claim 4; Fig 1A; 81pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid that
XX encodes a serine protease called D-G protein. The activity of the protein
XX of the invention may be described as cytostatic, antiinflammatory,
XX dermatological and anticoagulation. The serine protease of the invention
XX is a member of the trypsin/chymotrypsin-like (S1) serine protease family,
XX which play an important role in processes such as digestion and
XX regulatory amplification cascades through the proteolytic activation of
XX inactive zymogen precursors. Protease D-G modulating compounds are useful
XX for treating a condition which is mediated by protease D-G, e.g. cancer,

```


C skin disorders, neuropathic pain, inflammatory disorders, or coagulation
C diathesis/thrombosis. The polynucleotide encoding the protease is useful
C for identifying, detecting or isolating mutant forms of DNA molecules
C encoding the protease. The protease is useful for identifying modulators
C of the functional protease. The D-G protein can be used for formulation
C of compositions for laundry detergents and skin care products. Protease
C D-G gene therapy may be used to introduce protease D-G into the cells of
C target organisms. As the D-G protein is derived from a human, it is less
C likely to produce an allergic reaction in sensitive individuals when used
C in formulations for laundry detergents and skin care products. The
C current sequence represents the human serine protease D-G cDNA.

X
Q Sequence 2121 BP; 487 A; 614 C; 584 G; 436 T; 0 other;

Alignment Scores:

red. No.: 5.48e-209 Length: 2121
core: 2335.00 Matches: 434
percent Similarity: 99.77% Conservative: 0
est Local Similarity: 99.77% Mismatches: 1
uery Match: 99.70% Indels: 0
B: 24 Gaps: 0

S-09-607-745-2 (1-435) x AAI64284 (1-2121)

Y 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
b 277 ATGGATCCTGACAGTGAATCAACCTCTGAACAGCTCGATGTCAAAACCCCTGGCAACCC 336
Y 21 ArgIleProMetGluThrPheArgLysValGlyIleProIleIleLeuLeuLeuSer 40
b 337 CGTATCCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTAGC 396
Y 41 LeuAlaSerIleIleLeuValValLeuIleLysValIleLeuAspLysTyrTyrPhe 60
b 397 CTGGCGAGTATCATCATGTGTGTCTCATCAAGGTGATCTGGATAAATACTACTTC 456
Y 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
b 457 CTCTGGGGCAGCCTCTCCACTTTCATCCCGAGGAAGCAGCTGTGTGCGGAGAGCTGGAC 516
Y 81 CysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaVal 100
b 517 TGTCCCTTGGGGAGGAGGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGCGCTGCAGTG 576
Y 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
b 577 GCAGTCCGCTCTCCAGGACCGATCCACACTCGAGGTGTGGACTCGGCCACAGGGAAC 636
Y 121 TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
b 637 TGGTCTCTGCTGCTTCGACAACTTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAG 696
Y 141 MetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
b 697 ATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCAGACCATGATCTG 756
Y 161 AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180
b 757 GATGTTGTTGAATACAGAAACAGCCAGAGCTTCGCATGCGGAACCTCAAGTGGGCC 816
Y 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
b 817 TGTCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGGCTGTGGAGAGCCTGAAGACC 876
Y 201 ProArgValValGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
b 877 CCCCCTGTGGTGGGAGGAGGAGCCCTCTGTGGATTCTTGGCCTTGGCAGGTGAGCATC 936
Y 221 GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
b 937 CAGTACGACAAACAGCAGCTCTGTGGAGGAGCATCTCTGGACCCCACTGGTCTCTCAG 996
Y 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260

Db 997 GCAGCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGTGGCGAGGCTCA 1056
QY 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsn 280
Db 1057 GACAAACTGGGAGCTTCCCATCCCTGGCTGGCCCAAGATCATCATGAAATTCAC 1116
QY 281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
Db 1117 CCCATGTACCCCAAAGACAATGACATCGCCCTCATGAAGCTGCAGTCCCACACTTTC 1176
QY 301 SerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThr 320
Db 1177 TCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 1236
QY 321 ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle 340
Db 1237 CCACCTCTGGATCATTTGGATGGGCTTTACGAAGCAGAATGGAGGAAGATGTCTGACATA 1296
QY 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyr 360
Db 1297 CTGCTGCAGGCTCAGTCCAGGTCAATGACAGCACACGGTGAATGCAGACGATGCGTAC 1356
QY 361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr 380
Db 1357 CTGGGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGAAGGGGTGTGGACACC 1416
QY 381 CysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGly 400
Db 1417 TGCCAGGTGACAGTGGTGGGCCCTCTGATGTACCAATCTGCACAGTGGCATGTGGTGGC 1476
QY 401 IleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVal 420
Db 1477 ATCGTTAGCTGGGCTATGGCTGGGGGCCCTCTGATGTACCAATCTGCACAGTGGCATGTGGTGGC 1536
QY 421 SerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
Db 1537 TCAGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1581

RESULT 7
AAD02557
ID AAD02557 standard; DNA; 1479 BP.
XX
AC AAD02557;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human seripancrin variant #1 DNA.
XX
KW Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;
KW arthritis; chronic obstructive pulmonary disorder; COPD; cancer;
KW osteoporosis; aberrant wound healing; angiogenesis; diabetes;
KW inflammatory disorder; stroke; cardiovascular disease; gene therapy;
KW vaccine; cytostatic; cerebroprotective; vulnery; osteopathic; ds.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1479
FT /*tag= a
FT /product= "Human seripancrin variant #1 protein"
XX
PN WO200104141-A2.
XX
PD 18-JAN-2001.
XX
PF 04-JUL-2000; 2000WO-EP06211.
XX
PR 12-JUL-1999; 99EP-0113428.
XX
PA (MERE) MERCK PATENT GMBH.
XX
PI Suendermann B, Hofmann U, Matzku S, Wilbert O;

WPI; 2001-147177/15.
P-PSDB; AAY72559.

P-PSDB; AAY72559.

New extracellular serine protease Seripancrin, useful for treating cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease, diabetes, inflammatory disorders, stroke, angiogenesis and aberrant wound healing -

Disclosure; Page 40-42; 45pp; English.

The present invention relates to seripancrin polynucleotides, and polypeptides encoded by them. Seripancrin are members of serine protease family. This protein contains a transmembrane domain, a low density lipoprotein (LDL) domain, protease domain and a scavenger receptor cysteine-rich (SRCR) domain. The LDL and SRCR domains help to define the specificity of seripancrin's intra and intermolecular interactions. The polynucleotides and polypeptides of the invention are useful for treating and diagnosing diseases such as arthritis, chronic obstructive pulmonary disorder (COPD), cancer, osteoporosis, aberrant wound healing, angiogenesis, inflammatory disorders, diabetes, stroke and cardiovascular diseases. Seripancrin genes are useful in chromosome localisation studies, as tools for tissue expression studies and also in gene therapy. The polypeptides of the invention are used for identifying agonists and antagonists useful for treating conditions associated with seripancrin imbalance. These polypeptides are also useful as vaccines. The present sequence is a DNA coding for seripancrin variant #1 protein. The seripancrin gene is located on human chromosome 11q22-q23.

Sequence 1479 BP: 329 A; 425 C; 407 G; 318 T; 0 other;

ianment Scores:

Sequence ID:	1.24e-208	Length:	1479
Accession NO.:	2329.00	Matches:	432
Core:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Local Similarity:	100.00%	Indels:	0
Every Match:	99.44%	Gaps:	0
Identity:	22.4%		

1-09-607-745-2 (1-435) x AAD02557 (1-1479)

```

1  MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
  |||||
1  ATGGATCCTGACAGTGATCAACCTCTGAAACAGCCTCGATGTCAAACCCCTGCGCAAAACC 60
  |||||
21  ArgIleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSer 40
  |||||
61  CGTATCCCCATGGAGACCTTCAGAAAGGTGGGGATCCCCATCATAGCACACTAGTGGC 120
  |||||
41  LeuAlaSerIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPhe 60
  |||||
121  CTGGCGAGTATCATCATTTGTGGTTGTCTCATCAAGGTGATCTGGATAAATACTACTTTC 180
  |||||
61  LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
  |||||
181  CTCCTCGGGCAGCCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGGAGAGTGGAC 240
  |||||
81  CysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaVal 100
  |||||
241  TGTCCCTTGGSGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCAAGGGCCTGCAGTG 300
  |||||
101  AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
  |||||
301  GCAGTCCGCCCTCTCCAAGGACCGATCCACACTGCAGGTGCTGGACTCGGCCACAGGGAAAC 360
  |||||
121  TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
  |||||
361  TGGTTCTCTGCGCTGTTTCGACAACTTCCAGAAAGCTCTCGCTGACACAGCCTGTAGTCAG 420
  |||||
141  MetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
  |||||
421  ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGACCAGGATCTG 480
  |||||

```

XX
PN WO200274979-A2.
XX 26-SEP-2002.
XX
XX 20-MAR-2002; 2002WO-US08456.
XX 20-MAR-2001; 2001US-276947P.
XX
XX (ORTH) ORTHO CLINICAL DIAGNOSTICS INC.
XX Wan J, Wang Y;
XX WPI; 2002-740862/80.
XX
XX New gene expression profile generated from primary, endothelial,
XX epithelial, and muscle cell types, useful for identifying disease
XX pathologies involving alterations of gene expression, e.g. cancer -
XX
XX Disclosure; Page 759-760; 850pp; English.
XX
XX The invention relates to a gene expression profile comprising one or more
XX genes (ABZ34889-ABZ35692) and generated from a cell type. The cell type
XX is a coronary artery endothelium, umbilical artery or vein endothelium,
XX aortic endothelium, dermal microvascular endothelium, pulmonary artery
XX endothelium, myometrium microvascular endothelium, keratinocyte
XX epithelium, bronchial epithelium, mammary epithelium, prostate
XX epithelium, renal cortical epithelium, renal proximal tubule epithelium,
XX small airway epithelium, renal epithelium, umbilical artery smooth
XX muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,
XX dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,
XX aortic smooth muscle, mesangial cells, coronary artery smooth muscle,
XX bronchial smooth muscle, uterine smooth muscle, lung fibroblast,
XX osteoblasts or prostate stromal cell. The gene expression profile is used
XX for determining the level of RNA expression for a sample, determining the
XX phenotype of a cell and distinguishing cell types. The gene or a protein
XX expression profile is useful in identifying disease pathologies
XX involving alterations of gene expression. The assessment of expression
XX profiles may provide meaningful information with respect to tumour type
XX and stage, treatment methods, and prognosis. The gene or protein
XX expression profile may also be used for creating microarrays. The
XX microarray is useful for genetic and physical mapping of genomes, DNA
XX sequencing, genetic or medical diagnosis, genotyping of organisms,
XX confirming cell or tissue identifications and in identifying promising
XX antibiotics, antiviral or antifungal agents.
XX
XX Sequence 2165 BP; 502 A; 623 C; 595 G; 445 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 2,06e-208 Length: 2165
XX Score: 2329.00 Matches: 435
XX Percent Similarity: 99.77% Conservative: 0
XX Best Local Similarity: 99.77% Mismatches: 0
XX Query Match: 99.44% Indels: 1
XX DB: 24 Gaps: 0
XX
XX US-09-607-745-2 (1-435) x ABZ35523 (1-2165)
XX
XX 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
XX 310 ATGGATCCTGACAGTGATCACTCTGAACAGCCCTCATGTCAACCCCTGCGCAACCC 369
XX
XX 21 ArgIleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSer 40
XX 370 CGTATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 429
XX
XX 41 LeuAlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPhe 60
XX 430 CTGGGAGATATCATCTGTGTGTCTCATCAAGTGATTCGGATAAATACTACTTC 489
XX
XX 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
XX 490 CTCTCGGGGACGCTCTCCACTTCATCCCGAGGAACAGCTGTGTGACGGAGAGCTGGAC 549

QY 81 CysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaVal 100
Db 550 TGTCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCTGCAGTG 609
QY 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
Db 610 GCAGTCCGCCTCTCCAAAGGACCGATCCACACTGCAGGTGTGGACTCGGCACAGGGAAC 669
QY 121 TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
Db 670 TGGTCTCTGCCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAG 729
QY 141 MetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
Db 730 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCAGGATCTG 789
QY 161 AspValValGluIleThrGluAsnSerGlnGlu-LeuArgMetArgAsnSerSerGlyPr 180
Db 790 GATGTTGTTGAATCAAGAAAACAGCCAGGAGGCTTCGCATCGGGAATCAAGTGGGCC 849
QY 180 oCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysTh 200
Db 850 CTGTCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCCTGTGGGAAGACCTGAAGAC 909
QY 200 rProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIl 220
Db 910 CCCCCGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTTCAGCAT 969
QY 220 eGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuTh 240
Db 970 CCAGTACGACAAACAGCAGCTCTGTGGAGGAGCATCTGGACCCCTCTGGTCTCTCAC 1029
QY 240 rAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySe 260
Db 1030 GGACGCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGGTGGGCGGAGGCTC 1089
QY 260 rAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAs 280
Db 1090 AGACAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTAATTCAA 1149
QY 280 nProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPh 300
Db 1150 CCCCATGTACCCCAAGCAATGACATCGCCCTCATGAAGTGCAGTTCCTCACTCACTTT 1209
QY 300 eSerGlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaTh 320
Db 1210 CTCAGGCACAGTCAGGCCCATCTGTCTGCCCTCTTTGATGAGGAGCTCACTCCAGCCAC 1269
QY 320 rProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIl 340
Db 1270 CCCACTCTGGATCATGTGGTGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACAT 1329
QY 340 eLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTy 360
Db 1330 ACTGCTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCATGCAGACGATGCGTA 1389
QY 360 rGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspTh 380
Db 1390 CCAGGGGAAGTCAACGAGAAGATGATGTGTGAGGCATCCCCGAAGGGGTGTGGACAC 1449
QY 380 rCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGl 400
Db 1450 CTGCCAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACAGTGGCATGTGGTGGG 1509
QY 400 yIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVa 420
Db 1510 CATCGTTAGCTGGGCTATGGCTCGGGGGCCCCGAGCACCCCCAGGAGTATACACCAAGGT 1569
QY 420 lSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
Db 1570 CTCAGCCTATCTCAACTGGATCTACATATGTCTGGAAGGCTGAGCTG 1615

RESULT 9

AAZ90471

D AAZ90471 standard; cDNA; 2070 BP.

X C

C AAZ90471;

T X

T 06-JUN-2000 (first entry)

X X

E Cancer specific gene (CSG) sequence (clone ID 1283171).

X X

W CSG; cancer specific gene; cancer; gynecologic cancer; ovarian; breast;

W endometrial; uterine; lung; cytotoxic; ss.

X X

S Homo sapiens.

X X

N WO200012758-A1.

X X

D 09-MAR-2000.

X X

F 01-SEP-1999; 99WO-US19655.

X X

R 02-SEP-1998; 98US-0098880.

X X

A (DIAD-) DIADEXUS LLC.

X X

I Salceda S, Sun Y, Recipon H, Cafferkey R;

X X

R WPI; 2000-256657/22.

X X

T Diagnosing, staging, monitoring, imaging and treating cancer especially

T gynecological cancers e.g. breast, ovarian cancer and lung cancer,

T involves measuring cancer specific gene levels in cells and body fluids

X X

S Claim 9; Page 45-46; 58pp; English.

X X

The invention relates to detecting, diagnosing metastasis and staging cancer by measuring levels of cancer specific genes (CSG) in cells, tissues or body fluids. Their remission and progression, decreases and increases in CSG levels, is also monitored, by periodic sample analysis. The methods are useful for detecting cancers, especially gynecologic cancers which include ovarian, breast, endometrial and uterine cancer and lung cancer. Antibodies against the CSGs labeled with paramagnetic ions or a radioisotope is useful for imaging cancer and when conjugated with a cytotoxic agent are useful for treating cancer. The present sequence represents a CSG sequence (clone ID: 1283171 and gene ID: 332459).

X X

Sequence 2070 BP; 473 A; 599 C; 577 G; 421 T; 0 other;

X X

ignment Scores:

red. No.:

ore:

recent Similarity:

st Local Similarity:

ery Match:

;

-

-

-

-

-

-

-

-

-

-

-

-

-

-

-

-

-

-

403 TGCGGGCAGCCTCTCCACTTCCAGGAAAGCAGCTGTGTGACGAGAGCTGGACTGT 462
82 ProLeuGlyGluAspGluGluHisCysValysSerPheProGluGlyProAlaValAla 101
463 CCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCCTGCAGTGGCA 522
102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
523 GTCCGCCTCTCCAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGAACTGG 582
122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
583 TTCTCTGCCTGTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGCAGATG 642
142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
643 GGCTACAGCAGCAACCCACTTTCAGAGCTGTGGAGATTGGCCCCAGACAGGATCTGGAT 702
162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
703 GTTGTGAATCACAGAAACAGCCAGGAGCTTCGATGCGGAATCAAGTGGGCCCTGT 762
182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
763 CTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTGCTGCTGGGAGAGCCTGAAGACCCC 822
202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
823 CGTGTGTGGTGGGAGGAGGCTCTGTGGATCTTTGGCCTTGGCAGGTTCAGCATCCAG 882
222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr-Al 241
883 TACGACAAACAGCACGCTGTGTGGAGGAGCATCTGGACCCCTGAGTGGTCTCACGGGC 942
241 aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261
943 AGCCCACTGCTTCAGGAAACATACCGATGTGTCAACTGGAAGTCCGGGCAGGCTCAGA 1002
261 pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleLeuGluPheAsnPr 281
1003 CAAACTGGGCAGCTTCCCATCCCTGGTGTGGCAGGATCATCATGATGATTAATCAACC 1062
281 oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301
1063 CATGTACCCCAAGACAAATGACATGCGCCTCATGAAGTGCAGTTCCTCACTTTCTC 1122
301 rGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPr 321
1123 AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGTCACTCCAGCCACCCC 1182
321 oLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLe 341
1183 ACTCTGGATCATTTGATGGGCTTTACGAGCAGAAATGGAGGAAAGATGTCTGACATACT 1242
341 uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGl 361
1243 GCTGAGGCGTCAGTCCAGGTCATTGACAGACACAGGTGCAATGACAGCATGCGTACCA 1302
361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381
1303 GGGGGAAGTCACCGAGAAGATGATGTGTGAGGATCCCGGAAGGGGTGTGGACACCTG 1362
381 sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyI 401
1363 CCAGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGCAT 1422
401 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe 421
1423 CGTAGCTGGGCTATGGTGGGCGGCCAGCACCCAGGAGTATACACCAAGGCTTC 1482
421 rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435


```
|||||
b 1237 GCTCAGGGCGTCAGTCCAGGTCAATTGACAGACACAGGTGCAATGACAGCATGCGTACCA 1296
y 361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381
b 1297 GGGGGAAGTCACCGAGAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGACACTG 1356
y 381 sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyI1 401
b 1357 CCAGGTGACAGTGGTGGGCCCTGTATGATACCAATCTGACAGTGGCATGTGGTGGCAT 1416
y 401 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe 421
b 1417 CGTAGCTGGGCTATGGCTGCGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTC 1476
y 421 rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
b 1477 AGCCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1519
```

RESULT 11

AAI72975

AAI72975 standard; cDNA; 2079 BP.

AAI72975;

21-AUG-2002 (first entry)

CJA8 cDNA.

Gene; colorectal cancer; CGA7; CJA8; modulating protein; screening;
drug candidate; vaccine; ss.

Homo sapiens.

```
Key          Location/Qualifiers
CDS          302..1522
             /*tag= a
             /product= "CJA8"
```

US2002042067-A1.

11-APR-2002.

08-MAY-2001; 2001US-0851588.

17-AUG-2000; 2000US-0642252.
06-SEP-2000; 2000US-0656002.

```
(MACK/) MACK D.
(GISH/) GISH K C.
(WILS/) WILSON K E.
```

Mack D, Gish KC, Wilson KE;

WPI; 2002-453647/48.
P-PSDB; AAG79358.Screening drug candidates for treating colorectal cancer, comprises
determining the effect of the candidate on the expression profile gene
of CGA7 or CJA8 -

Disclosure; Fig 5; 40pp; English.

The sequences given in AAI72973-76 encode the colorectal cancer proteins, CGA7 and CJA8. The CGA7 and CJA8 proteins are colorectal cancer modulating proteins and have been mapped to chromosomes 2 (CGA7) and 11 (CJA8). These sequences may be used in the method of the invention for screening drug candidates. The method comprises adding a drug candidate to a cell that expresses an expression profile gene encoding CGA7, CJA8 or fragments and determining the effect of the drug candidate on the expression of the expression profile gene. The new methods are used to screen bioactive agents for the ability to bind to or modulate the activity of CGA7 or CJA8 and evaluate the effect of a

CC candidate colorectal cancer drug. An antibody to CGA7 or CJA8 can inhibit
CC the activity of CGA7 or CJA8, respectively, and is used to screen for
CC an agent that can interfere with the binding of CGA7 or CJA8 to the
CC antibody. The antibody can be used to treat colorectal cancer. The
CC antibody or a fragment of it is used to localize a therapeutic group to
CC a colorectal cancer tissue, where the therapeutic group is a cytotoxic
CC agent or a radioisotope. Antisense molecules are used to inhibit
CC colorectal cancer in a cell. Nucleic acid segments encoding CGA7 or
CC CJA8 are used in a biochip. CGA7, CJA8 or a nucleic acid encoding it
CC are used to elicit an immune response. CGA7 or CJA8 is used to
CC determine the prognosis of an individual with colorectal cancer.
CC Nucleic acid encoding CGA7 or CJA8 can be used in vaccines.

XX
SQ Sequence 2079 BP; 489 A; 594 C; 575 G; 421 T; 0 other;

Alignment Scores:

```
Pred. No.:      5.74e-208      Length:      2079
Score:          2324.00         Matches:      434
Percent Similarity: 99.77%      Conservative: 0
Best Local Similarity: 99.77%   Mismatches: 0
Query Match:     99.23%         Indels:      1
DB:              24             Gaps:        0
```

US-09-607-745-2 (1-435) x AAI72975 (1-2079)

QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21

Db 217 GATCCTGACAGTGTCAACCTCTGAACAGCCTCGATGTCAAACCCCTGCGCAACCCCGT 276

QY 22 IleProMetGluThrPheArgLys-ValGlyIleProIleIleIleAlaLeuLeuSerLe 41

Db 277 ATCCCCATGGAGACCTTCAGAAAGTGTGGGGATCCCCATCATCATAGCACTACTGAGCCT 336

QY 41 uAlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLe 61

Db 337 GCGAGTATCATATTGTGGTTGTCTCATCAAGGTGATTCTGGATAAATACTACTTCTCT 396

QY 61 uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCy 81

Db 397 CTGCGGGCAGCCTCTCCACTTTCATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGACTG 456

QY 81 sProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAl 101

Db 457 TCCCTTGGGGAGGAGGAGGAGCAGTGTCTCAAGAGCTTCCCGAAGGGCCTGCAGTGGC 516

QY 101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121

Db 517 AGTCCGCTCTCTCCAAGGACCGATCCACTGCAGGTGTGGACTCGGCCACAGGGAAGT 576

QY 121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141

Db 577 GTTCTCTGCTGTTCGACAACTTCACAGAAAGCTCTCGTGAGACAGCCTGTAGGCAGAT 636

QY 141 tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs 161

Db 637 GGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCAGACAGGATCTGGA 696

QY 161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181

Db 697 TGTGTGAAATCAGAAACACAGCCAGGAGCTTCGCATCGGAACTCAAGTGGGCCCTG 756

QY 181 sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr 201

Db 757 TCTCTCAGGCTCCCTGGTCTCCCTGCCTGTCTTGGCTGTGGAAAGAGCCTGAAGACCCC 816

QY 201 oArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleG1 221

Db 817 CCGTGTGGTGGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTTCAGCATCCA 876

QY 221 nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl 241

Db 877 GTACGACAAACAGCACGCTCTGTGGAGGAGGAGCATCTCTGGACCCCACTGGGTCTCTCAGGC 936

241 aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261
937 AGCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGGTGCGGCGAGGCTCAGA 996
261 pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPr 281
997 CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATCAATTCAACCC 1056
281 oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301
1057 CATGTACCCCAAGACATGACATGCCCTCATGAAGCTGCAGTTCCCACTCACTTCTC 1116
301 rGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPr 321
1117 AGGCACAGTCAGGCCCATCTCTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCC 1176
321 oLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLe 341
1177 ACTCTGGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGGAAGATGCTGACATACT 1236
341 uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGl 361
1237 CTTGCAGGCGTCAGTCCAGTCAATTGACAGCACACGGTGCAATGCAGACGATGCGTACCA 1296
361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381
1297 GGGGGAAGTCACCGAGAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGACACCTG 1356
381 sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIl 401
1357 CCAGGGTGACAGTGGTGGGCCCTGTATGTACCAATCTGACCAGTGGCATGTGGTGGCAT 1416
401 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe 421
1417 CGTTAGCTGGGCTATGCTGCGGGGCCGAGCACCCAGGAGTATACACCAAGGTCTC 1476
421 rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
1477 AGCCTATCTCAACTGGATCTACATATGCTGGAAGGCTGAGCTG 1519

ESULT 12

BS57763
D ABS57763 standard; cDNA; 2079 BP.
X C X
X ABS57763;
T 05-FEB-2003 (first entry)
X cDNA encoding human colorectal cancer modulating protein CJA8.
E Colorectal cancer modulating protein; BCMP; CJA8; colorectal cancer;
W gene therapy; human; gene; ss.

X Homo sapiens.

X Key Location/Qualifiers
T CDS 251..1522

T /*tag= a
T /product= "CJA8"
T /note= "Colorectal cancer modulating protein"
T /transl_except= (pos:314..316, aa:Leu)

X US6455668-B1.

X 24-SEP-2002.

X 06-SEP-2000; 2000US-0656002.

X 15-MAR-2000; 2000US-0525993.

X 28-JAN-2000; 2000US-0493444.

X (EOSB-) EOS BIOTECHNOLOGY INC.

PI Mack D, Gish KC, Wilson KE;
XX WPI; 2003-066245/06.
DR P-PSDB; ABG72428.
DR
XX Diagnosing colorectal cancer comprises determining the expression of a
PT gene encoding CJA8 in a first colon tissue of a first individual, and
PT comparing them to the expression of a gene encoding CJA8 in a second
PT normal tissue
XX
PS Claim 11; Fig 1; 31pp; English.
XX
CC The invention describes a method of diagnosing colorectal cancer
CC comprising determining the expression of a gene encoding CJA8 or its
CC fragment in a first colon tissue of an individual, and comparing the
CC expression of the gene to the expression of the gene in normal tissue
CC from the individual or in tissue from a second unaffected individual.
CC A difference in the expression indicates that the first individual has
CC colorectal cancer. The CJA8 genes and proteins are useful for diagnosing
CC and prognosticating colorectal cancer. The methods are also useful for
CC screening candidate bioactive agents that can modulate colorectal
CC cancer, or for treating or inhibiting colorectal cancer in a patient
CC e.g. by gene therapy. This sequence encodes the human colorectal
CC cancer modulating protein CJA8.
XX
SQ Sequence 2079 BP; 489 A; 594 C; 575 G; 421 T; 0 other;

Alignment Scores:
Pred. No.: 5.74e-208 Length: 2079
Score: 2324.00 Matches: 434
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.23% Indels: 1
DB: 25 Gaps: 0

US-09-607-745-2 (1-435) x ABS57763 (1-2079)

QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
Db 217 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAACCCCGT 276
QY 22 IleProMetGluThrPheArgLys-ValGlyIleProIleIleIleAlaLeuLeuSerLe 41
Db 277 ATCCCATGGAGACCTTCAGAAAGTGTGGGATCCCCATCATCATAGCACTACTAGCCT 336
QY 41 uAlaSerIleIleIleValValIleuIleLysValIleLeuAspLysTyrTyrPheLe 61
Db 337 GCGAGTATCATCATTTGTGTTGTCTCATCAAGGTGATTTGGATAAATACTACTTCTCT 396
QY 61 uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCy 81
Db 397 CTGCGGCGAGCCTCTCCACTTCATCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGACTG 456
QY 81 sProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaValAl 101
Db 457 TCCCTTGGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCGCTGCAGTGGC 516
QY 101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
Db 517 AGTCCGCTCTCCAAGGACCGATCCACACTGCAAGTGTGACTCGGCTCGGCCACAGGAACTG 576
QY 121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
Db 577 GTTCTCTGCTGTTTCGACAACTTCACAGAAGCTCTCGTGAGACAGCCTGTAGGCAGAT 636
QY 141 tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs 161
Db 637 GGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCAGACCCAGGATCTGA 696
QY 161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181
Db 697 TGTGTTGAAATCACAGAAAACACCCAGGAGCTTCGCATCGGGAACCTCAAGTGGGCCCTG 756

181	sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr	201
757	TCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTGGAGAGAGCTGAAGACCCC	816
201	oArgValValGlyGlyGluGluAlaSerValAspSerTirProTrpGlnValSerIleG1	221
817	CCGTGTGGTGGTGGGAGGAGGCCCTCTGTGGATTCTTGGCCCTTGGCAGTTCAGCATCCA	876
221	nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTirValLeuThrAl	241
877	GTACGACAAACAGCACGTCTGTGGAGGGAGCATCTCTGGACCCCACTGGGTCTCACGGC	936
241	aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs	261
937	AGCCCACTGCTTCAGGAACATACCGATGTGTCAACTGGAAGGTGCGGGCAGGCTCAGA	996
261	pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPr	281
997	CAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATATTGAATTCACCC	1056
281	oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe	301
1057	CATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCACCTTCTCTC	1116
301	rGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPr	321
1117	AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCC	1176
321	oLeuTirPilleIleGlyTirpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLe	341
1177	ACTCTGGATCATTTGGATGGGGCTTTACGAAGCAGATGGAGGGAAGATGTCTGACATACT	1236
341	uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrG1	361
1237	GCTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGCTGCAATGCAGACGATGCGTACCA	1296
361	nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy	381
1297	GGGGGAAGTCACCGAGAAGATGATGTGTGAGGCACTCCCGAAGGGGGTGTGGACACCTG	1356
381	sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIl	401
1357	CCAGGGTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGGCAT	1416
401	eValSerTirpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe	421
1417	CGTTAGTGGGCTATGGCTGCGGGGGCCCCGAGCACCCAGGAGTATACACCAAGGTCTC	1476
421	rAlaTyrLeuAsnTirPileTyrAsnValTirpLysAlaGluLeu	435
1477	AGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG	1519

RESULT 13

AD13114

AAD13114 standard; DNA; 2137 BP.

TTCTDHF

AAD13114:

LETTERS

16-OCT-2001 (first entry)

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Human membrane-type serine protease (MTSP) 3 DNA:

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

Human: transmembrane serine protease: membrane-type serine protease:

MTSP: protease domain: neoplastic disease: tumour: cancer: cytostatic:

lung carcinoma: colon adenocarcinoma: ovarian carcinoma: gene therapy:

lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy; ds

25.

Домо сари онс

Homo sapiens.

[illegible]

Key	Location/Qualifiers
cc. 1554	

261..1574
CDS

```
/*tag= a
```

```
/product= "Human transmembrane serine protease (MTSP)
```

XX	WO200157194-A2.
PN	
XX	
PD	09-AUG-2001.
XX	
PF	02-FEB-2001; 2001WO-US03471.
XX	
PR	03-FEB-2000; 2000US-0179982.
PR	18-FEB-2000; 2000US-0183542.
PR	22-JUN-2000; 2000US-0213124.
PR	26-JUL-2000; 2000US-0220970.
PR	08-SEP-2000; 2000US-0657986.
PR	22-SEP-2000; 2000US-0234840.
XX	
PA	(CORV-) CORVAS INT INC.
XX	
PI	Madison EL, Ong EO, Yeh J;
XX	
DR	WPI; 2001-488877/53.
DR	P-PSDB; AAE06931.
XX	
PT	Novel single chain polypeptide comprising protease domain of type-II
PT	membrane-type serine protease or its catalytically active portion
PT	useful for treating and preventing cancer and tumor -
XX	
PS	Claim 14; Page 197-199; 256pp; English.
XX	
CC	The invention relates to transmembrane serine proteases and their
CC	corresponding nucleotides and the protease domain of a type-II
CC	membrane-type serine protease (MTSP). MTSP is useful for identifying
CC	compounds that modulate or inhibits its proteolytic activity and for
CC	formulating a medicament for treating neoplastic disease. MTSP and
CC	its corresponding nucleotides are useful in preventing or treating
CC	tumours or cancers such as lung carcinoma, colon adenocarcinoma and
CC	ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP
CC	is useful as a diagnostic marker for tumour development, growth and/or
CC	progression and as immunogens to generate antibodies that specifically
CC	bind to it. MTSP DNA is useful in a yeast two-hybrid system and in
CC	gene therapy. The present sequence is a DNA encoding human MTSP3
CC	protein.
XX	
SQ	Sequence 2137 BP; 508 A; 612 C; 589 G; 428 T; 0 other;
	Alignment Scores:
	Pred. No.: 1.75e-207 Length: 2137
	Score: 2319.00 Matches: 432
	Percent Similarity: 99.54% Conservative: 0
	Best Local Similarity: 99.54% Mismatches: 2
	Query Match: 99.02% Indels: 0
	DB: 22 Gaps: 0
	US-09-607-745-2 (1-435) x AAD13114 (1-2137)
QY	2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
Db	270 GATCCTGACAGTGATCAACCTCTGAAACAGCCTCGATGTCAAACCCCTCGCGAACCCCGT 329
QY	22 IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeu 41
Db	330 ATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACACTAGCAGCCTG 389
QY	42 AlaSerIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
Db	390 GCGAGTATCATCTTGTGGTGTCTTCATCAAGGTGATTCTGGATAAATACTACTTCCTC 449
QY	62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
Db	450 TGCGGGCAGCCCTCTCCACTTCATCCCGAGGAAGCAGTGTGTGACGGAGAGCTGGACTGT 509
QY	82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
Db	510 CCCTTTGGGGGAGGACGAGGAGACTGTGTCAAGAGCTTCCCGAAGGGCCTGCAGTGCGCA 569

102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
|||||
570 GTCCGCTCTCCAAGACCGATCCACACTGCAGTGTGGACTCGGCCACAGGGAAGTGG 629

122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
|||||
630 TTCTCTGCCTGTTTCGACAACTTCACAGAAGCTCTCGTGTAGACAGCCTGTAGGCAGATG 689

142 GlyTyrSerSerLysProThrPheArgAlaValGluLeuGlyProAspGlnAspLeuAsp 161
|||||
690 GGCTACAGCAGCAAAACCCACCTTCAGAGCTGTGGAGATTGGCCACAGACCAGGATCTGGAT 749

162 ValValGluLeuThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
|||||
750 GTTGTGAATCACAGAAAACAGCCAGGAGCTTCGATGCGGAACCTCAAGTGGGCCCTGT 809

182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
|||||
810 CTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGGCTGTGGGAAGAGCCTGAAGACCCCC 869

202 ArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
|||||
870 CGTGTGGTGGTGGGAGGAGGCCCTCTGTGGATTCTTGGCCTTGGCAGGTCCAGTCCAG 929

222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
|||||
930 TAGCACATACAGCACGCTCTGTGGAGGGAGCATCTCGACCCCACTGGTCTCTCACGGCA 989

242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
|||||
990 GCCCACTGCTTCAGGAACATACCGATGTGTTCACTGGAAGTGGGGCAGGCTCAGAC 1049

262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
|||||
1050 AAAGTGGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATCAATCAATCAACCCC 1109

282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
|||||
1110 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGTGCAGTTCCTCACTCACTTCTCA 1169

302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
|||||
1170 GGCACAGTCAGGCTCATCTGTCTGCCCTTCTTTGATGAGGAGTCACTCAGCCACCCCA 1229

322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
|||||
1230 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGAAGATGTCTGACATCTG 1289

342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 361
|||||
1290 CTGCAGGCGTCAGTCCAGTCAATGACAGCACACGTCGAATGCAGACGATGCGTACCAG 1349

362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
|||||
1350 GGGGAAGTCACCGAGAAGATGATGTGTGAGGATCCCGGAAGGGGGTGTGGACACCTGC 1409

382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
|||||
1410 CAGGGTGACAGTGGTGGGCCCTGATGTATCCATCTGACCAGTGGCATGTGGTGGGCATC 1469

402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValThrLysValSer 421
|||||
1470 GTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCAGGAGTATACCAAGGTCTCA 1529

422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
|||||
1530 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGTGAGCTG 1571

ESULT 14
AAA37099

D AAA37099 standard; cDNA; 2063 BP.

X

C. AAA37099;

XX 08-AUG-2000 (first entry)
DT Human PRO1570 (UNQ776) cDNA sequence SEQ ID NO:274.
XX
DE Human PRO1570 (UNQ776) cDNA sequence SEQ ID NO:274.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
KW ss.
XX Homo sapiens.
OS
PN WO200012708-A2.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.

R 30-SEP-1998; 98US-0102484.
R 30-SEP-1998; 98US-0102487.
R 30-SEP-1998; 98US-0102570.
R 30-SEP-1998; 98US-0102571.
R 01-OCT-1998; 98US-0102684.
R 01-OCT-1998; 98US-0102687.
R 02-OCT-1998; 98US-0102965.
R 06-OCT-1998; 98US-0103258.
R 06-OCT-1998; 98US-0103449.
R 07-OCT-1998; 98US-0103314.
R 07-OCT-1998; 98US-0103315.
R 07-OCT-1998; 98US-0103328.
R 07-OCT-1998; 98US-0103395.
R 07-OCT-1998; 98US-0103396.
R 07-OCT-1998; 98US-0103401.
R 08-OCT-1998; 98US-0103633.
R 08-OCT-1998; 98US-0103678.
R 08-OCT-1998; 98US-0103679.
R 08-OCT-1998; 98US-0103711.
R 14-OCT-1998; 98US-0104257.
R 20-OCT-1998; 98US-0104987.
R 20-OCT-1998; 98US-0105000.
R 20-OCT-1998; 98US-0105002.
R 21-OCT-1998; 98US-0105104.
R 22-OCT-1998; 98US-0105169.
R 22-OCT-1998; 98US-0105266.
R 26-OCT-1998; 98US-0105693.
R 26-OCT-1998; 98US-0105694.
R 27-OCT-1998; 98US-0105807.
R 27-OCT-1998; 98US-0105881.
R 27-OCT-1998; 98US-0105882.
R 27-OCT-1998; 98US-0106062.
R 28-OCT-1998; 98US-0106023.
R 28-OCT-1998; 98US-0106029.
R 28-OCT-1998; 98US-0106030.
R 28-OCT-1998; 98US-0106032.
R 28-OCT-1998; 98US-0106033.
R 28-OCT-1998; 98US-0106178.
R 29-OCT-1998; 98US-0106248.
R 29-OCT-1998; 98US-0106384.
R 29-OCT-1998; 98US-0108500.
R 30-OCT-1998; 98US-0106464.
R 03-NOV-1998; 98US-0106856.
R 03-NOV-1998; 98US-0106902.
R 03-NOV-1998; 98US-0106905.
R 03-NOV-1998; 98US-0106919.
R 03-NOV-1998; 98US-0106932.
R 03-NOV-1998; 98US-0106934.
R 10-NOV-1998; 98US-0107783.
R 17-NOV-1998; 98US-0108775.
R 17-NOV-1998; 98US-0108779.
R 17-NOV-1998; 98US-0108787.
R 17-NOV-1998; 98US-0108788.
R 17-NOV-1998; 98US-0108801.
R 17-NOV-1998; 98US-0108802.
R 17-NOV-1998; 98US-0108806.
R 17-NOV-1998; 98US-0108807.
R 17-NOV-1998; 98US-0108867.
R 17-NOV-1998; 98US-0108925.
R 18-NOV-1998; 98US-0108848.
R 18-NOV-1998; 98US-0108849.
R 18-NOV-1998; 98US-0108850.
R 18-NOV-1998; 98US-0108851.
R 18-NOV-1998; 98US-0108852.
R 18-NOV-1998; 98US-0108858.
R 18-NOV-1998; 98US-0108904.

(GETH) GENENTECH INC.

Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;

WPI; 2000-237871/20.

P-PSDB; AAY99417.

XX New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
PT small molecule inhibitors of the relevant receptor/ligand interactions
XX Claim 2; Fig 155; 773pp; English.
PS
XX
CC AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC transmembrane and receptor PRO proteins can be used for screening of
CC potential peptide or small molecule inhibitors of the relevant
CC receptor/ligand interactions. The polypeptides and nucleotide sequences
CC encoding then have various industrial applications, including uses as
CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC PCR primers and hybridisation probes used in the isolation of the PRO
CC polypeptides from the present invention.
XX
SQ Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

Alignment Scores:
Pred. No.: 1.73e-205 Length: 2063
Score: 2297.50 Matches: 429
Percent Similarity: 98.85% Conservative: 0
Best Local Similarity: 98.85% Mismatches: 0
Query Match: 98.10% Indels: 5
DB: 21 Gaps: 1

US-09-607-745-2 (1-435) x AAA37099 (1-2063)

QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
Db 219 GATCCTGACAGTGATCAACCTCTGAACAGCTCGATGTCAAAACCTGGCAAACCCCGT 278
QY 22 IleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSerLeu 41
Db 279 ATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATATAGACACTACTGAGCCTG 338
QY 42 AlaserIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
Db 339 GCGAGTATCATATTGTGTTGTCTCATCAAGGTGATTCTGGATAAATACTACTTCCTC 398
QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
Db 399 TGGGGCAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGAGGAGAGCTGGACTGT 458
QY 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
Db 459 CCCTTGGGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGAGCTGCAGTGGCA 518
QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
Db 519 GTCCGCCTCTCCAAGGACCGATCCACACTGCGAGGTGTGACTCGGCCACAGGAACTGG 578
QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
Db 579 TTCTCTGCTGTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGCAGATG 638
QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
Db 639 GGCTACAGC-----AGAGCTGTGGAGATTGGCCCCAGACAGGATCTGGAT 683
QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
Db 684 GTTGTGAAATCACAGAAACAGCCAGGAGTTCGCATGCGGAACCTCAAGTGGGCCCTGT 743
QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
Db 744 CTCACAGGCTCCCTGGTCTCCCTGCACGTGTCTGTGCTGTGGGAGAGCCTGAAGACCCC 803
QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
Db 804 CGTGTGGTGGTGGGGAGGAGGCCTCTGTGTGATTCTTGGCCTTGGCAGGTGAGCATCCAG 863

222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
864 TACGACAAACAGCAGCTGTGTGGAGGAGCATCTGGACCCCACTGGGTCTCACGGCA 923
242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
924 GCCCACTGCTTCAGGAACATACCGATGTGTCAACTGGAAGGTGGGGCAGGTCAGAC 983
262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleLeuGluPheAsnPro 281
984 AAACCTGGGCAGCTTCCCATCCCTGGCTGGCTGGCCCAAGATCATCATATTGAATTCACCC 1043
282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
1044 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCCCACTCACTTTCTCA 1103
302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163
322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
1164 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACTG 1223
342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 361
1224 CTGCAGCGTCAGTCCAGGTCATTGACAGCACAGCGTGCATGCAGACGATCGGTACCCAG 1283
362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
1284 GGGGAAGTCACCGAGAAGATGATGTGTCAGGCATCCCGAAGGGGTGTGGACACCTGC 1343
382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
1344 CAGGGTGACAGTGGTGGSCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGCATC 1403
402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
1404 GTTAGCTGGGCTATGGCTGCGGGGGCCCGACCCAGGAGTATACCAAGGTCTCA 1463
422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
1464 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505

RESULT 15

AS46089

D AAS46089 standard; cDNA; 2063 BP.

X X X X X

C C C C C

X T T T T

X X X X X

Human DNA encoding PRO polypeptide sequence #165.

PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
PCR primer.

Homo sapiens.

WO200168848-A2.

X N N N X

X X X X X

D D D D D

X X X X X

28-FEB-2001; 2001WO-US06520.

F F F F F

X X X X X

R R R R R

R R R R R

R R R R R

R R R R R

R R R R R

PR 14-MAR-2000; 2000US-189328P.
PR 15-MAR-2000; 2000WO-US06884.
PR 21-MAR-2000; 2000US-190828P.
PR 21-MAR-2000; 2000US-191007P.
PR 21-MAR-2000; 2000US-191048P.
PR 21-MAR-2000; 2000US-191314P.
PR 28-MAR-2000; 2000US-192655P.
PR 29-MAR-2000; 2000US-193032P.
PR 29-MAR-2000; 2000US-193053P.
PR 30-MAR-2000; 2000WO-US08439.
PR 04-APR-2000; 2000US-194449P.
PR 04-APR-2000; 2000US-194647P.
PR 11-APR-2000; 2000US-195975P.
PR 11-APR-2000; 2000US-196000P.
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199550P.
PR 25-APR-2000; 2000US-199654P.
PR 03-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 05-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-0644848.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.

(GETH) GENENTECH INC.

Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;

Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;

WPI; 2001-602746/68.

P-PSDB; AAU29188.

Novel nucleic acids encoding PRO polypeptides, used to diagnose the presence of tumours, such as prostate and breast tumours, in mammals and to screen for modulators of the compounds -

Claim 2; Fig 329; 774pp; English.

Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR primers for PRO polypeptides of the invention. The sequences of the invention can be used to detect the presence of a tumour in a mammal by comparing the level of expression of a PRO polypeptide in a test sample of cells from the animal and a control sample of normal cells, whereby a higher level of expression in the test sample indicates the presence of a tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats and rabbits but are preferably human. The polypeptides can be used to stimulate tumour necrosis factor (TNF) alpha release from human blood, when contacted with it. A specific polypeptide can be used to stimulate the proliferation or differentiation of chondrocyte cells. The PRO proteins can be used to determine the presence of tumours and also susceptibility to tumour development, particularly adrenal, lung, colon, breast, prostate, rectal, cervical, or liver tumours, in mammalian subjects. The oligonucleotide probes specific for the PRO nucleic acids can be used for genetic analysis of individuals with genetic disorders.

Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

Alignment Scores:

Pred. No.:

Score: 1.73e-205

Percent Similarity: 2297.50

Best Local Similarity: 98.85%

Length: 2063

Matches: 429

Conservative: 0

Mismatches: 0

every Match:	98.10%	Indels:	5
:	22	Gaps:	1
-09-607-745-2 (1-435) x AAS46089 (1-2063)			
2	AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg	21	
219	GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAACCCCTGGCGAAACCCCGT	278	
22	IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeu	41	
279	ATCCCCATGGAGACCTTCAGAAAGGTGGGATCCCCATCATCATAGCACTACTGAGCCTG	338	
42	AlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu	61	
339	GGGAGTATCATCATTTGTGGTTGTCTCATCAAGGTGATTCTGGATAAATACTACTTCCTC	398	
62	CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys	81	
399	TGCGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGACTGT	458	
82	ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla	101	
459	CCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGCCTGCAGTGGCA	518	
102	ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp	121	
519	GTCCGCCTCTCCAAGGACCGATCCACACTGCAGGTGCTGGACTCGGCCACAGGAACTGG	578	
122	PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet	141	
579	TTCTCTGCCTGTTTCGACAACATTACAGAAAGCTCTCGCTGAGACAGCCTGTAGGCAGATG	638	
142	GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp	161	
639	GGCTACAGC-----AGAGCTGTGGAGATTGGCCACAGCCAGGATCTGGAT	683	
162	ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys	181	
684	GTTGTTGAATCACAGAAAACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCCCTGT	743	
182	LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro	201	
744	CTCTCAGGCTCCCTGGTCTCCCTGCACGTGTTGCCTGTGGGAAGAGCCTGAAGACCCCC	803	
202	ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln	221	
804	CGTGTGTGGTGGGAGGAGGCCTCTGTGGATTCTTGGCCTTGGCAGGTTCAGCATCCAG	863	
222	TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla	241	
864	TACGACAAACAGCACGCTCTGTGGAGGGAGCATCCTGGACCCCCACTGGGTCTTCACGGCA	923	
242	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	261	
924	GCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCAGAC	983	
262	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro	281	
984	AAACTGGGCAGCTTCCCCATCCCTGGCTGTGGCCAAAGATCATCATTAATTC AACCCC	1043	
282	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	301	
1044	ATGTACCCCAAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCACACTCACTTTCTCA	1103	
302	GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	321	
1104	GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA	1163	
322	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu	341	
1164	CTCTGGATCATTTGATGGGGCTTTACGAAGCAGAAATGGAGGGGAAGATGTCTGACATACTG	1223	

QY	342	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	361
Db	1224	CTGCAGGGCTCAGTCCAGGTCAATTGACAGCACACGGTGCAATGCAGACGATCGGTACCAG	1283
QY	362	GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	381
Db	1284	GGGGAAGTCACCGAGAAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGACACCTGC	1343
QY	382	GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	401
Db	1344	CAGGTGACAGTGGTGGSCCCTGATGTATACCAATCTGACCAGTGGCATGTGTGGGCATC	1403
QY	402	ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer	421
Db	1404	GTTAGCTGGGGCTATGGCTGCGGGGCCCCGAGCACCCCCAGGAGTATACACCAAGGTCTCA	1463
QY	422	AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	435
Db	1464	GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG	1505

Search completed: November 23, 2003, 08:06:43
Job time : 440.919 secs